INTER-AMERICAN DEVELOPMENT BANK
BANCO INTERAMERICANO DE DESARROLLO (BID)
RESEARCH DEPARTMENT
COMPETITIVENESS STUDIES SERIES
WORKING PAPER C-104

DESIGNING COMPETITIVE WHOLESALE ELECTRICITY MARKETS FOR LATIN AMERICAN COUNTRIES

FRANK A. WOLAK

STANFORD UNIVERSITY

Prepared for the First Meeting of the Latin American Competition Forum,
a joint endeavour of the Organisation for Economic
Co-Operation and Development (OECD) and IDB
Paris, France
April 7-8, 2003

The author gratefully acknowledges the collaboration of the Sustainable Development Department (SDS), particularly Jaime Millán.

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Wolak, Frank A.

Designing competitive wholesale electricity markets for Latin American countries / Frank A. Wolak.

p. cm.
(Competitiveness Studies Series working papers ; C-104)
Includes bibliographical references.

1. Competition—Latin America. 2. Electric utilities—Latin America. 3. Electric utilities--Deregulation—Latin America. I. Inter-American Development Bank. Research Dept. II. Title. III. Series.

333.7932 W593--------dc21

The Research Department (RES) produces a quarterly newsletter, *IDEA (Ideas for Development in the Americas)*, as well as working papers and books on diverse economic issues. To obtain a complete list of RES publications, and read or download them please visit our web site at: http://www.iadb.org/res

Executive Summary

The purpose of this paper is to present a general framework for electricity market design in Latin American Countries (LACs) that addresses the current problems facing electricity supply industries (ESIs) in this region. The major issue addressed is what market rules, market structures, and legal and regulatory institutions are necessary to establish a competitive wholesale market that provides the maximum possible benefits to consumers consistent with the long-term financial viability of the ESI.

The paper first presents a theoretical foundation for analyzing the electricity market design problem. A generic principal-agent model is presented and its applicability to the electricity market design problem explained. It is then applied to illustrate the incentives for firm behavior under regulation versus market environments. The impact of government versus private ownership on firm behavior in both market and regulated environments is also addressed using this model. This discussion is used to guide our choices for the important lessons for electricity market design in developed countries and LACs.

Using the experiences from ESI reform in developed countries, the paper presents five essential features of a successful wholesale electricity market. The first is the need for a sufficient number of independent suppliers for a competitive market to be possible. Merely declaring the market open to competition will not result in new entry unless no single supplier is able to dominate the market. Second is a forward market for electricity where privately-owned firms are able to sell long-term commitments to supply electricity. This report argues that the conventional wisdom of establishing a competitive spot market first leading to a competitive forward market is an extremely expensive process in developed countries and is prohibitively expensive in developing countries. Third is the need for the active involvement of as many consumers of electricity as is economically feasible in the operation of the wholesale market. This involvement should occur both in the long-term and short-term market. In the short-term market, there must be a number of buyers willing to alter their consumption of electricity in response to short-term price signals. Fourth is the importance of a transmission network to facilitate commerce, meaning that the transmission network must have sufficient capacity so that all suppliers face significant competition. This implies a dramatically different approach to determining the quantity and magnitude of transmission network expansions in a market regime. The final lesson is the need to establish a credible regulatory mechanism as early as possible in the restructuring process. An important lesson from developed countries around the world is that the initial market design will have flaws. This implies the need for ongoing market monitoring to correct these flaws before they develop into disasters.

The paper then takes on the issue of the specific challenges to LAC restructuring. Rather than focus on the details of specific markets, the paper instead identifies a number of problems common to LACs and provides recommended solutions to each of these problems. A major theme of this section is a warning that short-term solutions to market design flaws can have long-term market efficiency costs. The paper identifies seven major challenges to Latin American ESI restructuring. The first is related to the problem of introducing wholesale markets in systems dominated by hydroelectric capacity. This section also deals with the related issue of how using cheap hydroelectric power keeps electricity prices low but the risk of electricity shortages high. The second issue is concerned with the difficulties of establishing an active forward market for electricity in LACs. The third relates to the LAC-specific challenges associated with establishing an independent and regulatory body. The fourth addresses the advisability of cost-based versus bid-based dispatch of generation units in LAC wholesale markets. The fifth is how to regulate the default provider retail electricity price in LACs. The sixth concerns the advisability of capacity payments mechanism for ensuring energy adequacy in markets where demand is expected to grow rapidly. The final issue is the role of government versus private ownership in LACs.

The report then discusses specific market design challenges in five LACs. These countries are Brazil, Chile, Colombia, Honduras, and Mexico. A number of these challenges are specific examples of the general challenges discussed earlier in the paper, whereas others are unique to the geography, natural resource base or legal environment in the country.

The report closes with a proposed market design that should serve as a baseline market design for all LACs. Deviations from this basic design could be substantial depending on initial conditions in the industry and the country, but the ideal behind proposing this design is to have a useful starting point for all LAC restructuring processes.

Contents

1. Introduction

One lesson from the past decade of electricity supply industry (ESI) reforms in developed countries is that it is easy to make extremely costly mistakes and very difficult to avoid making any mistakes. Wholesale market meltdowns have occurred in California and in New Zealand. All other countries with wholesale electricity markets have experienced sustained periods of very high prices that cannot be explained by fuel or other input cost increases; many of these episodes have overwhelmed the regulatory institution that oversees the wholesale market. These experiences suggest that the only thing a government considering ESI reform can be sure of is that mistakes will be made in the initial market design and that the regulatory institution will face a number of difficult challenges that it is not prepared to deal with.

Because the legal and regulatory institutions necessary for a competitive wholesale electricity market are less mature in Latin American Countries (LACs) than in developed countries, successful power sector reform should be even more difficult in these countries. Further complicating this task is the fact that the pre-reform financial condition of the ESIs in most LACs is much worse than the pre-reform financial health of the ESIs in all of the developed countries that have restructured. The primary motivation for restructuring in LACs is to attract private investment capital to all aspects of the industry in a fast-growing market with a limited transmission and distribution network infrastructure. In contrast, the primary motivation for restructuring in developed countries is to improve the operating efficiency of existing capacity and impose greater economic discipline on new capacity investment decisions in slower-growing markets with more than adequate generation capacity and extensive transmission and distribution networks. These divergent motivations make restructuring a higher-risk activity in LACs. However, if done correctly, restructuring is also a higher expected return activity because of its more ambitious goals in LACs. The paper proposes a generic restructuring process for LACs that limits these risks so that significant potential net benefits of ESI restructuring can be realized both in the short run and long run.

The purpose of this paper is to present a framework for analyzing the electricity market design process in LACs that addresses the problems facing all ESIs, as well as those that are unique to countries in this region. The major issue addressed is: What market rules, market structures, and legal and regulatory institutions are necessary to establish competitive wholesale markets that provide the maximum possible benefits to consumers consistent with the long-term financial viability of ESIs in LACs?

This paper will first use outcomes from the past decade of ESI restructuring in the United States, Europe and Australia and New Zealand to identify the major challenges facing all electricity market design processes. These challenges will then be placed in the LAC context. For example, the extreme dependence of many LACs on hydroelectric power makes the experience of hydro-dependent countries such as Norway and New Zealand particularly instructive. The paper will then identify the wholesale electricity market design problems unique to LACs. These problems and their importance to ESI reform in LACs will be identified with respect to the market structure and performance of the ESIs in five LACs that I visited as background for preparing this report: Brazil, Chile, Colombia, Honduras and Mexico. The countries further along in this process—Brazil, Chile and Colombia—are used to identify problems that arise in the operation of a wholesale market and its regulatory oversight in LACs. The experiences of all five countries provide useful background for my discussion of the impediments to the reform process in LACs.

My analysis is organized around five major themes. First is the necessity of a forward market for electricity where privately owned firms are able to sell long-term commitments to supply electricity. I will argue that the conventional wisdom of first establishing a competitive spot market in order to foster an active forward market has proven to be an extremely expensive undertaking in developed countries. I believe it is prohibitively expensive in most developing countries. Experience from developed country ESI restructuring around the world has shown that the major source of supply-side benefits from industry restructuring is the competitive procurement of long-term energy commitments of sufficient magnitude and duration to allow suppliers to fund the construction of new generation facilities. The spread of wholesale forward markets throughout the US during the early 1980s led to new generation capacity investment decisions driven by purely economic factors. As a result, the vast majority of new generation facilities have significantly lower operating costs than existing generation capacity. This paper suggests a market design process for LACs that maximizes the likelihood that a transparent and active forward market for electricity will form.

A second theme deals with the difficulties created by restructuring an ESI with significant dependence on hydroelectric capacity. All LACs with significant hydroelectric capacity have experienced energy shortages relative to demand at the prevailing retail price before, during, or after the transition to a wholesale market regime.[1] Different from fossil fuel-based electricity systems, higher electricity prices do not increase the supply of the input energy source. Specifically, rainfall does not increase in response to higher electricity prices. This implies that the greater the share of electricity produced using hydroelectric capacity the larger is the potential risk of these events. Because most existing hydroelectric capacity in LACs was constructed during the former state-owned monopoly regime, governments find it difficult to resist increasing this risk of shortages by overusing water during low rainfall years. This leads to artificially low wholesale electricity prices that discourage investment in fossil-fuel sources that would provide much-needed insurance against water shortages. This paper suggests a number of market rule, market structure and regulatory oversight changes to limit the incentives for inefficient use of water and increase the incentives for new investment in more reliable electricity sources.

The third theme emphasizes and elaborates on the need for the active involvement of as many electricity customers as is economically feasible in the operation of the wholesale market. Final customers must be actively involved in both the long-term and short-term market. In the long-term market, wholesale electricity purchasers, typically retailers and large industrial and commercial consumers, must be permitted to make the long-term purchasing commitments with electricity suppliers that allow these firms to finance generation capacity expansion. In the short-term market, there should be as many final electricity consumers as possible able to benefit from altering their hourly electricity consumption in response to short-term price signals. Flexibility in consumption is crucial to increasing the competitiveness of the short-term energy and ancillary services market. The ability to shift significant amounts of electricity demand across hours of the day or days of the week in response to price changes is far more important to limiting the market power of firms than is the ability to reduce consumption during all hours of the day in response to price changes.

---

[1] See Fischer and Galetovic (2001) for the case of Chile, Ayala and Millan (2002) for the case of Colombia, and Parente (2002) for the case of Brazil.

The fourth theme is the importance of adequate transmission capacity to support a competitive wholesale market. Here I will make the distinction between a reliable transmission network to support a competitive wholesale market and a reliable transmission network to support a vertically integrated monopoly ESI. Because of the initial condition of the transmission network in most LACs, I will argue that this distinction has important implications for the design of the wholesale market and the role of the independent system operator.

The final theme is often overlooked and its importance the most underappreciated in the market design process. This is the need for a credible regulatory mechanism as early as possible in the restructuring process. Spiller and Martorell (1996) argue that this is a key factor in the success of ESI reform in Chile. As discussed above, all governments and regulators can be virtually certain that the initial wholesale market design will have flaws. This implies the need for ongoing market monitoring to correct these flaws before they develop into disasters. As the California electricity crisis demonstrates, even the most experienced regulator can make enormous mistakes. Even though regulation is an imperfect process, it is most effective when the regulator has a reputation for technical and economic expertise and independence from the government in power. Consequently, perhaps the most important challenge facing regulators in LACs is how to establish a reputation for technical and economic expertise so that as few decisions as possible are overturned on judicial review or even taken to judicial review. I will provide recommendations for solving this initial conditions problem in establishing a credible regulatory process that protects consumers from the exercise of market power and suppliers from the attempts of politicians to expropriate the value of their investments.

2. The Market Design Problem

There are two primary dimensions of the market design problem. The first is the extent to which market mechanisms versus regulatory processes are used to set the prices consumers pay. The second is the extent to which market participants are government versus privately owned. Given the technologies for producing and delivering electricity to final consumers in a country, the market designer faces two basic challenges. The first is how to cause producers to supply electricity in both a technically and allocatively efficient manner. Technically efficient production obtains the maximum amount of electricity for a given quantity of inputs, such as capital, labor, materials and input energy. Allocatively efficient production uses the minimum cost mix of inputs to produce a given level of output.

The second challenge is how to cause the simultaneous actions of all suppliers and retailers to set the lowest possible retail price consistent with the long-term financial viability of the industry. Consequently, the goal of the market design process is to devise mechanisms for compensating market participants for their actions, so that final consumers pay the lowest possible retail prices needed for the industry to sustain itself over the long-term. This involves choosing a point in the market versus regulation dimension and government versus private ownership dimension for each segment of the electricity supply industry.

Conceptually, the market designer chooses the number and sizes of each market participant and the rules for determining the revenues received by each market participant to maximize its objective function. There are two key constraints on the market designer's optimization problem. The first is that once the market designer selects the rules for determining the revenues each market participant receives, the market participant will choose a strategy that maximizes his payoff given the rules set by the market designer. This constraint implies that the market designer must recognize that all market participants will maximize expected profits given the rules the market designer selects. The second constraint is that each market participant must expect to receive from the compensation scheme chosen by the market designer more than their opportunity cost of participating in the ESI. The first constraint is called the individual rationality constraint because it assumes each market participant will behave in a rational (expected payoff-maximizing) manner. The second constraint is called the participation constraint, because it implies that firms must expect that participation in the ESI is more attractive than their next best alternative.

2.1. The Principal-Agent Problem

To make these features of the market design problem more concrete, it is useful to consider a very special case of this process—the generic principal-agent model. Here a single principal designs a compensation scheme for a single agent that maximizes the principal's expected payoff, subject to the agent's individual rationality constraint and a participation constraint. Let $W(x,s)$ denote the payoff of the principal given the observable outcome of the interaction, x, and state of the world, s. The observable outcome, x, depends on the agent's action, a, and the true state of the world. In general, x is written as $x(a,s)$ to denote the fact that it depends on both of these variables.

Let $V(a,y,s)$ equal the payoff of the agent when it takes the action, a, faces the compensation scheme set by the principal, $y(x)$, and true the state of the world is s. The principal's action is to design the compensation scheme, $y(x)$, a function that relates the outcome observed by the principal, x, to the payment made to the agent.

With this notation, it is possible to define the two constraints facing the principal in designing $y(x)$. The individual rationality constraint on the agent's behavior is that it will choose its action, a, to maximize its payoff $V(a,y,s)$ (or the expected value of this payoff) given $y(x)$ and s (or the distribution of s). The participation constraint implies that the compensation scheme $y(x)$ set by the principal must allow the agent to achieve at least its reservation level of utility or expected utility V^*. There are two basic versions of this model. The first assumes that the agent does not observe the true state of the world when it takes its action, and the other assumes the agent observes s before taking its action. In the first case, the agent's choice is:

$$a^* = \underset{a}{\operatorname{argmax}} E_s(V(a, y(x), s)),$$

where $E_s(.)$ denotes the expectation with respect to the distribution of s. The participation constraint is $E_s(V(a^* ,y(x^*),s)) > V^*$, where $x^* = x(a^*,s)$. In the second case, the agent's problem is:

$$a^*(s) = \underset{a}{\operatorname{argmax}} V(a, y(x), s),$$

and the participation constraint is $V(a^*(s) ,y(x^*),s) > V^*$ for all s, where $x^* = x(a^*(s),s)$ in this case.

An enormous number of bilateral economic interactions fit this generic principal-agent framework. Examples include the client-lawyer, patient-doctor, lender-borrower, employer-worker, and firm owner-manager interactions. A client seeking legal services designs a compensation scheme for her lawyer that depends on the observable outcomes (such as the verdict in the case) that causes the lawyer to maximize the client's payoff function subject to the constraint that the lawyer will take actions to maximize his payoff given this compensation scheme and the fact that the lawyer must find the compensation scheme sufficiently attractive to take on the case. Another example is the firm owner designing a compensation scheme that causes the manager to maximize the value of the owner's assets subject to the constraint that the firm manager will take actions to maximize her payoff given the scheme is in place and the fact that it must provide a higher payoff to the manager than she could receive elsewhere.

2.2. Applying the Principal-Agent Model to the Market Design Process

An example of this principal-agent model relevant to ESI restructuring is the regulator-utility interaction. In this case, the regulator designs a scheme for compensating the vertically integrated monopoly for the actions that it takes, recognizing the fact the once this regulatory mechanism is put in place the utility will attempt to maximize its payoff function given this regulatory mechanism. In this case, $y(x)$ would be the mechanism used by the regulator to compensate the firm for its actions. For example, under a simple *ex post* cost-of-service regulatory mechanism, x would be the output produced by the firm, and $y(x)$ would be the firm's total cost of providing this output. Under a price cap regulatory mechanism, x would be the change in the consumer price index for the United States (US) economy and $y(x)$ would be the total revenues the firm receives, assuming it serves all demand at the price set by this regulatory mechanism. The incentives for firm behavior created by any potential regulatory mechanism can be studied within the context of this principal-agent model.

This modeling framework is also useful for understanding the incentives for firm behavior in a market environment. A competitive market is another possible way to compensate a firm for the actions that it takes. For example, the regulator could require this firm and other firms to bid their willingness to supply as a function of price and only chose the firms with bids below the lowest price necessary to meet the aggregate demand for the product. In this case x

can be thought of as the firm's output and *y(x)* the firm's total revenues from producing *x* and being paid this market-clearing price per unit sold. Viewed from this perspective, markets are simply another regulatory mechanism for compensating a firm for the actions it takes.

It is well known that profit-maximizing firms participating in a competitive market have a strong incentive to produce their output in a technically and allocatively efficient manner. However, it is also well known that profit-maximizing firms have no unilateral incentive to pass on these minimum production costs in the price they charge to consumers. It is only when competition among firms is sufficiently fierce that this will occur.

Economic theory provides conditions under which a market will yield an optimal solution to the problem of causing the suppliers to provide their output to consumers at the lowest possible price. One of these conditions is the requirement that suppliers are atomistic, meaning that all producers believe they are so small relative to the market that they have no ability to influence the market price through their actions. Unfortunately, this condition is unlikely to hold for the case of electricity given the size of most market participants before the reform process starts. These firms recognize that if they remain large, they will have the ability to influence both market and political outcomes through their unilateral actions. Moreover, the minimum efficient scale of electricity generation, transmission and distribution is such that it is unlikely to be least costly for the industry as a whole to separate electricity production into a large number of extremely small firms. So there is an underlying economic justification for allowing these firms to remain large, although certainly not as large as they would like to be. This is one reason why the electricity market design process is so difficult. This problem is particularly acute for small Latin American countries such as Honduras and other small Central American countries without substantial transmission interconnections with neighboring countries. The minimum efficient scale for a new combined cycle gas turbine (CCGT) facility may be a substantial fraction of the peak demand in these countries.

This principal-agent model is also useful for understanding why industry outcomes can differ so dramatically depending on whether the industry is government or privately-owned. First, the objective function of the firm's owner differs across the two regimes. Under government ownership all of the citizens of the country are shareholders. These owners are also severely limited in the sorts of mechanisms they can design to compensate the management of the firm. For example, there is no liquid market for selling their ownership stake in this firm. It is virtually impossible for them to remove the management of this firm. They don't even have a legal right to their ownership stake in the firm. In contrast, a shareholder in a privately owned firm has a clearly defined and legally enforceable property right that can be sold in a relatively liquid market. If they own enough shares in the firm or can get together with other large shareholders, they can remove the management of the company. Finally, by selling their shares, they can severely limit the ability of the company to raise capital for new investment, because a lower stock price means that the management must now issue more shares to raise the same amount of money. In contrast, the government-owned firm obtains the funds necessary for new investment primarily through the political process, so there is very little a private citizen can do to prevent a politically savvy government-owned firm from undertaking new investment.

This discussion illustrates the point that despite the fact that both the government-owned and privately-owned firm have access to exactly the same technologies to generate, transmit and distribute electricity, dramatically different industry outcomes in terms of the mix of generation capacity installed, the price consumers pay and the amount they consume can occur because the schemes for compensating each firm's management, $y(x)$, differ. Because the owners of the two firms have different objective functions and face different sets of feasible mechanisms for compensating their management, they find it optimal to select different mechanisms for compensating the management for the actions it takes.

Applying the generic principal-agent model to the issue of government versus private ownership implies that different industry outcomes should occur if a government-owned, vertically integrated geographic monopolist is asked to provide electricity to the same geographic area that a privately owned geographic monopolist previously served, even if both monopolists face the same regulatory mechanism for setting the prices they charge to retail consumers.

Applying the logic of the principal-agent model at the level of the regulator-firm interaction, as opposed to the firm owner-management interaction, implies an additional source of differences in market outcomes if, as is often the case, the government-owned monopoly faces a different regulatory process than the privately owned monopoly.

In the competitive market context, the extent of government participation in the industry creates an additional source of differences in industry outcomes. Because the nature of the principal-agent relationship between the firm's owner and its management is different under private ownership versus government ownership, an otherwise identical government-owned firm can be expected to behave differently in a market environment from how this firm would behave if it were privately owned. Therefore, even for two markets with the same number and sizes of firms, market outcomes will differ depending on the ownership status (government owned versus privately owned) of all of firms in the market. These differences in market outcomes can be particularly significant if the largest firms in the industry are government owned instead of privately owned.

Consequently, in its most general form, the market design problem is composed of multiple layers of principal-agent interactions where the same principal can often interact with a number of agents. For example, in the case of a competitive wholesale electricity market, the same regulator interacts with all the firms in the ESI. The market designer must recognize the impact of all of these principal-agent relationships in designing an electricity supply industry to achieve her market design goals. The vast majority of electricity market design failures result from ignoring the individual rationality constraints implied by both the regulator-firm and firm owner-management principal-agent relations. The individual rationality constraint most often ignored is that privately-owned firms will maximize their profits from participating in a wholesale electricity market given the market rules set by the market designers. In addition, if offered the opportunity to provide input to the market design process, individual rationality would also imply that market participants would argue for market rules that enhance their ability to maximize expected profits.

It is important to emphasize that this individual rationality constraint holds whether or not the privately owned profit-maximizing firm is one of a number of firms in a market environment or a single vertically integrated monopolist. The only difference between these two environments is the set of actions that the firm is legally able to take to maximize its profits.

2.3. Individual Rationality Under a Market Mechanism versus a Regulatory Process

The set of actions available to firms in a market environment generally differ from those available to it in a regulated-monopoly environment. For example, under a market mechanism firms can increase their profits by both reducing the costs of producing a given level of output or by increasing the price they charge for this output. In contrast, in a regulated monopoly environment, the firm does not set the price it receives for its output. Instead, the legal contract between the firm and regulator usually requires the firm to supply all that is demanded at a price set by the regulator in exchange for the firm being given a legal monopoly to supply all customers in a given geographic area at the price set by the regulator, as well as the opportunity to earn a reasonable rate of return on their investment from the prudent operation of their facilities.

Defining the incentive constraint for a privately owned firm operating in a competitive electricity market is relatively straightforward. Because the firm would like to maximize profits, it has a strong incentive to produce any amount of output at minimum cost. In other words, the firm will find it individually rational to produce in a technically and allocatively efficient manner. However, as discussed above, the firm has little incentive to set a price that only recovers these production costs. In fact, the firm would like to take actions to raise the price it receives above both the average and marginal cost of producing its output. Profit-maximizing behavior implies that the firm will choose a price or level of output such that the increase in revenue it earns from supplying one more unit equals the additional cost that it incurs from producing one more unit of output. This is the same thing as saying that the firm will withhold output from the market until the cost savings from withholding one more unit of output is less than or equal to the total revenue loss from withholding that unit of output from the market. This is a more general restatement of a standard result from profit-maximizing behavior when the firm has the ability to influence the market price: The firm will produce at the point that the marginal revenue from selling an additional unit of output equals the marginal cost of producing that unit of output.

Figure 1 provides a simple model of the unilateral profit-maximizing behavior of a supplier in a bid-based electricity market. Let Q_d equal the level of market demand for a given hour and $SO(p)$ the aggregate willingness to supply as a function of the price for all other market participants besides the firm under consideration. Figure 1(a) plots the inelastic aggregate demand curve and the upward sloping supply curve for all other firms besides the one under consideration. Figure 1(b) subtracts this aggregate supply curve for all other market participants from the market demand to produce the residual demand curve faced by this supplier, $DR(p)$ =

Q_d - SO(p). This panel also plots the marginal cost curve for this supplier, as well as the marginal revenue curve associated with *DR(p)*. The intersection of this marginal revenue curve with the supplier's marginal cost curve yields the profit-maximizing level of output and market price for this supplier given the bids submitted by all other market participants. This price-quantity pair is denoted by *(P^*,Q^*)* in Figure 1(b). Profit-maximizing behavior by the firm implies the following relationship between the marginal cost at Q^*, which I denote by *$MC(Q^*)$*, and P^* and „ the elasticity of the residual demand at P^*:

$$(P^* - MC(Q^*))/P^* = - 1/„ \qquad (1)$$

where *, = DRN(P^*)*(P^*/DR(P^*))*. Because the slope of the firm's residual demand at this level of output is finite, the market price is larger than the supplier's marginal cost. The price-quantity pair associated with the intersection of *DR(p)* with the supplier's marginal cost curve is denoted *(P^c,Q^c)*. It is important to emphasize that even though the price-quantity pair *(P^c,Q^c)* is often called the competitive output level, producing at this level is not unilateral profit-maximizing for the firm if it faces a downward-sloping residual demand curve. This is another way of saying that price-taking behavior—the firm acting as if it had no ability to impact the market price—is never individually rational. It will only occur as an equilibrium outcome if competitive conditions in the market are particularly fierce.

Figure 1(a)-(b) illustrates the essential difference between the firm's unilateral profit-maximizing level of output when it has the ability to influence the market price through its own actions and its profit-maximizing level output when the supplier believes it has no ability to influence the market price because of competitive conditions in the market. The supplier withholds output from the market relative to the price-taking outcome, *(P^c,Q^c)*, because it knows that by doing so, it raises the price that it receives for all the units it sells. The firm continues to withhold output from the market until the cost reduction from withholding an additional unit of output from the market is exactly balanced by the revenue reduction the supplier experiences from withdrawing that unit of output from the market. In contrast, a price-taking firm produces at the point that marginal cost associated with an additional unit of output is equal to consumers' willingness to pay for that additional unit of output.

A firm that influences market prices as shown in Figure 1(a)-(b) is said to be exercising its unilateral market power. A firm possesses unilateral market power if it has the ability to raise the market price through its unilateral actions and profit from this price increase. We would expect all privately owned profit-maximizing firms to behave in this manner. This is equivalent to saying that the firm satisfies its individual rationality constraint. I would like to emphasize that as long as a supplier faces a residual demand curve with any upward slope, it has the ability to exercise unilateral market power.

Figure 1(c)-(d) illustrates the extremely unlikely case that the supplier faces an infinitely elastic residual demand curve and therefore finds it in its unilateral profit-maximizing to produce at the point that the market price is equal to its marginal cost. This point is denoted (P^{**},Q^{**}). The supplier faces an infinitely elastic residual demand curve because the *SO(p)* curve is infinity elastic at P^{**}, meaning that all other firms besides this supplier are able to produce all that is demanded if the price is above P^{**}. Note that even in this extreme case the supplier is still producing at the point where the marginal revenue curve associated with *DR(p)* crosses its marginal cost curve. The only difference is that this marginal revenue curve is also equal to its average revenue curve, because *DR(p)* is infinitely price elastic, meaning that it is a horizontal line. Even in this extreme case, the firm continues to set prices that satisfy equation (1). However, because the slope of the firm's residual demand curve is infinite, 1/„ and is equal to zero so that equation (1) implies producing at the point where price equals marginal cost.

Figure 1 demonstrates that the individual rationality constraint in the context of a market mechanism is equivalent to the supplier exercising all available unilateral market power. Even in the extreme case of an infinitely elastic residual demand curve given Figures 1(c) and 1(d), the supplier still exercises all available unilateral market power. However, in this case the supplier cannot increase its profits by withholding output that can be produced at a marginal cost less than market price, because the firm possesses no unilateral market power, which means that it is unable to raise market price by its own actions.

Individual rationality in the context of a regulatory process still implies that the firm will maximize profits given the mechanism for compensating it set by the regulator. However, in this case the firm is usually unable to set the price it charges consumers or the level of output it supplies because of the explicit or implicit regulatory contract described above. Consequently, the firm must take more subtle approaches to maximizing its profits because the regulator typically sets the output price and requires the firm to supply all that is demanded at this regulated price. In this case the individual rationality constraint can imply that the firm will produce its output in a technically or allocatively inefficient manner because of how the regulatory process sets the price that the firm is able to charge. For example, the well-known Averch and Johnson (1962) model of cost-of-service regulation assumes that the regulated firm produces its output using capital, K, and labor, L, yet the price the regulator allows the firm to charge for capital services in setting its output price is greater than the actual price the regulated firm pays for capital services. This implies that a profit-maximizing firm facing the pricing-setting constraint implied by this regulatory process will produce its output using capital more intensively relative to labor than would be the case if the regulatory process did not allow a different price for capital services from the one the firm actually pays. The Averch and Johnson model illustrates a very general point associated with the individual rationality constraint in regulated settings: It is virtually impossible to design a regulatory mechanism that causes a privately owned profit-maximizing firm to produce in a least-cost manner.

The major reason why the regulator is unable to set prices that achieve the market designer's goal of least-cost production is that the regulated firm usually knows more about its production process or demand than the regulator. Although both the firm and regulator have substantial expertise in the technology of generating, transmitting and distributing electricity to final consumers, the firm has a much better idea of precisely how these technologies are implemented to serve its customers. This informational asymmetry leads to disputes between the firm and the regulator over the minimum cost mode of production to serve the firm's demand. Consequently, the regulator can never know the minimum cost mode production to serve final demand.

Moreover, there are laws against the regulator confiscating the firm's assets through the prices it sets, and the firm is aware of this fact. This creates the potential for disputes between the firm and the regulator over the price level that provides strong incentives for least-cost production but does not confiscate the firm's assets. All governments recognize this fact and allow the firm the opportunity to subject the regulator to judicial review of any regulatory decision about the level of the firm's price. To avoid the expense and potential loss of credibility of a judicial review, the regulator may instead prefer to set a slightly higher regulated price to guarantee that the firm will not submit this decision for judicial review. This informational asymmetry in the regulatory process reduces the firm's incentive to produce its output in a least-cost manner.

Wolak (1994) studies the regulator-utility interaction between California water utilities and the California Public Utilities Commission. He specifies and estimates an econometric model of this principal-agent interaction and quantifies the magnitude of the distortions from minimum cost production induced by the informational asymmetry between firm and the regulator about one aspect of the firm's production process. Even for the very straightforward technology of providing local water delivery services, where the extent of informational asymmetries between the firm and the regulator is likely to be small, Wolak (1994) finds that actual production costs are between 5 percent and 10 percent higher than they would be under least-cost production. This result suggests that the deviations from least-cost production in the electricity supply industry are likely to be much greater because the extent of the informational asymmetries between the firm and regulator about the firm's production process are much greater.

The market designer need not worry about the impact of informational asymmetries on a firm's mode of production in a competitive market. There is no legal requirement that a market mechanism set the price a firm receives above some minimum level. Different from price-regulated environments, there are no laws against a competitive market setting prices that confiscate a firm's assets. Any firm that is unable to cover its costs of production at the market price must eventually exit the industry. Firms cannot file for judicial review of the prices set by a competitive market, unless they believe some antitrust law has been violated. Competition among firms leads high-cost firms to exit the industry and be replaced by lower-cost firms. Contrary to the regulated regime, there is no need to determine if a firm's incurred production costs are the result of a least-cost mode of production. If the market is sufficiently competitive and has low barriers to entry, then any firm that is able to remain in business must be producing its output at or close to minimum cost. Otherwise, a more efficient firm could enter and profitably undercut the price of this firm. The risk that firms not producing in a least-cost manner will be forced to exit creates much stronger incentives for least-cost production than would be the case under regulation, where the firm recognizes that the regulator does not know the least-cost mode of production and that it can exploit this fact through technically and allocatively inefficient production that may ultimately yield the firm higher profits.

This difference in the incentives for least-cost production under regulation versus a market mechanism reinforces the impact of individual rationality constraints on firm behavior under a competitive market regime versus a regulated monopoly regime. In the case of a market mechanism the individual rationality constraint (to maximize expected profits) provides strong incentives for each firm to produce its output at least cost, but little, if any, incentive to price this output to only recover production costs. In fact, depending on the extent of competition the firm faces, it may have an extremely strong incentive to price its output vastly in excess of the marginal cost of producing the most expensive unit sold.

For the case of the regulated monopoly regime, the individual rationality constraint implies that the firm does not produce its output in a least-cost manner. However, because the regulator sets the price the firm is able to charge, this price only recovers the firm's incurred costs. Consequently, the advantage of regulation is that the market price should not deviate significantly from the average cost of producing the firm's output. However, the firm has very little incentive to make its actual mode of production equal to the least-cost mode of production. In contrast, the market regime provides very strong incentives for firms to produce in a least-cost manner but—unless the market is competitive—little incentive to pass on these low production costs in the prices charged to consumers.

This discussion shows that the potential exists for consumers to pay lower prices under either regime. Regulation may be favored if the market designer is able to implement a regulatory process that is particularly effective at causing the firm to produce in a least-cost manner, or if the market designer is unable to establish a sufficiently competitive market so that prices are vastly in excess of the marginal cost of producing the last unit sold. Competition is favored if regulation is particularly ineffective at providing incentives for least-cost production or if competition is particularly fierce. Nevertheless, in making the choice between market mechanisms and regulatory mechanisms the market designer must make a choice between two imperfect worlds. Which mechanism should be selected depends on which one maximizes the market designer's objective function.

2.4. Individual Rationality Constraint Under Government versus Private Ownership

The individual rationality constraint for a government-owned firm is difficult to characterize for two reasons. First, it is unclear what control the firm's owners are able to exercise over the firm's management and employees. Second, it is also unclear what the objective function of the firm's owners is. For the case of privately owned firms, there are well-defined answers to both of these questions. The firm's owners have clearly specified legal rights, and their ownership shares can be bought and sold by incurring modest transaction costs. Because, keeping all other things equal, investors would like to earn the highest possible return on their investments, and shareholders would like the firm's management to maximize the risk-adjusted rate of return on equity. This implies that the firm's owners will attempt to devise a compensation scheme for the firm's management that causes them to maximize expected profits. By comparison, it is unclear if the government wants its firms to maximize expected profits. Earning more revenues than costs is clearly a priority, but once this is accomplished the government would most likely want the firm to pursue other goals.

This lack of clarity in both the objective function of the government for the firms it owns and the set of feasible mechanisms the government can implement to compensate the firm's management has a number of implications. The first is that it is unlikely that the management of a government-owned firm will produce and sell its output in a profit-maximizing manner. Different from a privately-owned firm, its owners are not demanding the highest possible return on their equity investments in the firm. However, because a government-owned firm's management has little incentive to maximize profits, it also has little incentive to produce in a least-cost manner. By the same token, this logic also implies that a government-owned firm has little incentive to attempt to raise prices beyond the level necessary to cover its total costs of production. The second implication of this lack of clarity in objectives and feasible mechanisms is that the firm's management now has the flexibility to pursue a number of other goals besides minimizing the total cost of producing the output demanded by consumers.

Viewed from the perspective of the overall market design problem, one advantage of government-ownership is that the pricing goals of the firm do not directly contradict the market designer's goal of the lowest possible prices consistent with the long-term financial viability of the industry. In the case of private ownership, the pricing incentives of the firm's management directly contradict the interests of consumers. As discussed in the previous section, the firm's management wants to raise prices above the marginal cost of the last unit produced, because of the desire of the firm's owners to receive the highest possible return on their investment in the company. The desire of privately owned firms to maximize expected profits leads to pricing incentives that directly contradict the goals of the market design process. Unless the firm faces sufficient competition from other suppliers, which from the discussion of Figure 1 is equivalent to saying that the firm faces a sufficiently elastic residual demand curve, this desire to raise the market price will yield market outcomes that cause significant harm to consumers.

However, it is important to emphasize that prices set by a government-owned firm may cause at least as much harm to consumers as prices that reflect the exercise of unilateral market power if the incentives for least-cost production by the government-owned firm are sufficiently muted and the firm is required to set a price that at least recovers all of its incurred production costs. Although these prices may appear more benign because they only recover the actual costs incurred by the government-owned firm, they are in fact more harmful from a societal welfare perspective than the same level of prices set by a privately owned firm. This is because the privately owned firm has a strong incentive to produce in a technically and allocatively efficient manner and any positive difference between total revenues paid by consumers and the minimum cost of producing the output sold is economic profit or producer surplus. However, for the case of the government-owned firm there is another reason why the firm is required to raise its price. That is because it is producing in a technically and allocatively inefficient manner, which is socially wasteful and therefore yields a reduced level of producer surplus relative to the case of a privately owned firm. Because both outcomes, by assumption, have consumers paying the same price, the level of consumer surplus is unchanged across the two ownership structures, so that the level of total surplus is reduced as a result of government ownership.

Figure 2 illustrates this point. The step function labeled MC_p is the incurred marginal cost curve for the privately owned firm and the step function labeled MC_g is the incurred marginal cost curve for the government-owned firm. I make the distinction between incurred and minimum cost to account for the fact that the management of the government-owned firm has less of an incentive to produce at minimum cost than does the privately owned firm. In this example, I assume the reason for this difference in marginal cost curves is that the government-owned firm uses more units of each input to produce the same level of output than does the privately owned firm. Suppose that the profit-maximizing level of output for the privately owned firm, given the residual demand curve plotted in Figure 2, is Q^*, with a price of P^*. Suppose the government-owned firm behaves as if it were a price-taker given its marginal cost curve and this residual demand curve, and assume that this price is also equal to the firm's average incurred cost at Q^*, AC(Q^*). I have drawn the figure so that the intersection of the marginal cost curve of the government-owned firm with this residual demand curve occurs at the same price and quantity pair. However, as noted above, the government-owned firm uses much more of society's scarce resources to produce Q^* than the privately owned firm. Consequently, the additional benefit that society receives from having the privately owned firm produce the good, even though it is exercising significant unilateral market power, is the shaded area between the two marginal cost curves in Figure 2. This example demonstrates that even though the privately owned firm exercises all available unilateral market power, if the incentives for efficient production by government-owned firms are sufficiently muted, it may be preferable from the market designer's and society's perspective to tolerate some exercise of unilateral market power, rather than adopt a regime with government-owned firms setting prices equal to an extremely inefficiently incurred marginal cost or average cost of production.

The example given in Figure 2 may seem extreme, but there are a number of reasons why it is reasonable to believe that a government-owned firm faces far less pressure from its owners to produce in a least-cost manner than its privately owned counterpart. For example, poorly run privately-owned companies can go bankrupt. If a firm's creditors are not paid, they can demand to have the firm liquidate its assets to pay them. If a firm consistently earns revenues less than its production costs, the firm's owners and creditors will force the firm to liquidate its assets and exit the industry. The experience from both developed and developing countries is that poorly run government-owned companies rarely go out of business. Governments can and almost always do fund unprofitable companies from general tax revenues. Even in the United States, there are a number of examples of persistently unprofitable government-owned companies receiving subsidies long after the time that the vast majority of independent observers say that these firms should liquidate their assets and exit the industry. Because government-owned companies have this additional source of funds to cover their incurred production costs, they have significantly less incentive to produce in a least-cost manner.

3. Lessons for Designing a Competitive Wholesale Market from Developed Countries

Although there have been some highly publicized wholesale electricity market design failures in developed countries, there have also been a number of wholesale market design successes from ESI restructuring processes over the past fifteen years. This section will describe the important positive lessons from these reforms. This involves describing five essential initial conditions necessary to have a competitive wholesale electricity market. Because countries have and will continue to implement wholesale markets without these initial conditions in place, I will describe a number of safeguards that limit the potential harm to consumers from implementing ESI reforms with less-than-optimal initial conditions. I will also discuss the long-term implications of these safeguards, because many of them provide short-term protection but hinder long-term market efficiency.

3.1. Essential Features of a Competitive Wholesale Market

As has been emphasized by a number of observers, spot electricity markets are extremely susceptible to the exercise of unilateral market power. Borenstein, Bushnell and Wolak (2002, hereafter BBW) present estimates of the extent of unilateral market power exercised in the California electricity market over the period June 1998 to October 2000. Joskow and Kahn (2002) perform a similar analysis that focuses on the events of the summer of 2000 in the California market and the issue of withdrawing capacity from the market to exercise unilateral market power. Bushnell, Mansur, and Saravia (2002) compare the extent of unilateral market power exercised in the California market to that in the PJM (portions of Pennsylvania, New Jersey and Maryland as well as Delaware and Washington, D.C.) and ISO (Independent System Operator)-New England wholesale markets. The major conclusion from this three-market analysis is that unilateral market power is common to all of these wholesale markets, particularly during system conditions when the demand for electricity is sufficiently high that a large fraction of the within-control-area generating capacity is needed to meet this demand.

As discussed in Section 2, it is impossible to eliminate the incentive that suppliers in a competitive electricity market have to exercise unilateral market power. The best that a market designer can hope to do is reduce it. Using the framework of Section 2, this means the market designer must recognize the individual rationality constraint that the firm will maximize profits given the market rules and actions taken by the firm's competitors. As the discussion of Figure 1 demonstrates, the market designer reduces the incentive the firm has to exercise unilateral market power by facing the firm with a residual demand curve that is as elastic as possible. Although I do not expect the firm's desire to maximize profits to be diminished by facing it with a more elastic residual demand curve, as Figure 1 demonstrates, the more elastic the supplier's residual demand curve, the less the firm's unilateral profit-maximizing actions are able to raise the market-clearing price. Consequently, the goal of designing a competitive electricity market is straightforward: Face all suppliers with as elastic as possible residual demand curves during as many hours of the year as possible.

Wolak (2003b) presents evidence consistent with this market design goal. Using bid data from the California Independent System Operators's (CAISO) real-time electricity market, he computes, ϵ_{jh}, the elasticity of the hourly residual demand curve for hour h facing supplier j evaluated at the hourly market-clearing price for each of the five large in-state suppliers to the California electricity market—AES/Williams, Duke, Dynegy, Mirant and Reliant—for the period June 1 to September 30 for 1998, 1999 and 2000. Consistent with the market-wide estimates of the extent of unilateral market power exercised presented in BBW, Wolak (2003b) demonstrates that for all of these suppliers the average hourly value of $1/\epsilon_{jh}$ was higher in 2000 relative to 1998 and 1999. This result implies that the ability of each of these five suppliers to raise market prices by bidding to maximize their profits from selling electricity in the CAISO's real-time market was much greater in 2000 relative to the previous two years. The average hourly value of $1/\epsilon_{jh}$ in 1998 was somewhat higher than the same value in 1999, indicating the unilateral profit-maximizing actions of suppliers to the California market in 1999 were less able to raise market prices than in 1998. This result is consistent with the market-wide estimates of the extent of unilateral market power exercised in BBW for 1998 versus 1999.

There are five primary mechanisms for increasing the elasticity of the residual demand curve faced by a supplier in a wholesale electricity market. The first is divestiture of capacity owned by this firm into a larger number of independent suppliers. Second is the magnitude and distribution across suppliers of financial forward contracts to supply electricity to load-serving entities. Third is the extent to which final consumers are active participants in the wholesale electricity market. Fourth is the extent to which the transmission network has sufficient capacity to deliver electricity to all locations in the transmission network so that each firm faces sufficient competition from other suppliers. The last is the extent to which regulatory oversight of the wholesale market provides strong incentives for all market participants to fulfill their contractual obligations and obey the market rules. We now discuss each of these mechanisms for increasing the elasticity of the residual demand curve facing a supplier.

3.1.1. Divestiture of Suppliers

To understand how the divestiture of a given amount of capacity into a larger number of independent suppliers can impact the slope of the residual demand a firm faces, consider the following simple example. Suppose there are ten equal-sized firms, each of which owns 1,000 MW of capacity and that the total demand in the hourly wholesale market is equal to 9,500 MWh. Each firm knows that at least 500 MW of its capacity is needed to meet this demand, regardless of the actions of its competitors. Specifically, if the remaining nine firms bid all 1,000 MW of their capacity into the market, the tenth firm has a residual demand of at least 500 MWh at every bid price. Mathematically, this means the value of the residual demand facing the firm, $DR(p)$, is positive at p_{max}, the highest possible bid price that a supplier can submit. When $DR(p_{max}) > 0$, the firm is said to be pivotal, meaning that at least $DR(p_{max})$ of its capacity is needed to serve demand. Figure 3 provides an example of this phenomenon. Let $SO_1(p)$ represent the bid supply curve of all other firms besides the firm under consideration and Q_d the level of demand. Figure 3(b) shows that the firm is pivotal for $DR_1(p_{max})$ units of output, which in this example is equal to 500 MWh. In this circumstance, the firm is guaranteed total revenues of at least $DR_1(p_{max})*p_{max}$, which it can achieve by bidding all of its capacity in at p_{max}.

To see the impact on a firm's residual demand curve from requiring it to sell capacity, suppose that the firm in Figure 3 was forced to sell off 500 MW of its capacity to a new entrant to the market. This implies that the maximum supply of all other firms is now equal to 9,500 MWh, the original 9,000 MWh plus the additional 500 MWh divested, which is exactly equal to the level of demand. This means that the firm is no longer pivotal, because its residual demand is equal to zero at p_{max}. Figure 3(a) draws a new bid supply curve of all other market participants besides the firm under consideration, $SO_2(p)$. For every price, I would expect this curve to lie to the right of $SO_1(p)$, the original bid supply curve. Figure 3(b) plots the resulting residual demand curve for the firm using $SO_2(p)$. This residual demand curve, $DR_2(p)$, crosses the vertical axis at p_{max}, so that the elasticity of the residual demand curve facing the firm is now finite for all feasible prices. In contrast, for the case of $DR_1(p)$, the residual demand pre-divestiture, the firm faces an inelastic demand of at least $DR_1(p_{max})$ for all prices in the neighborhood of p_{max}.

This is an example of a general phenomenon associated with structural divestiture: the firm now faces a more elastic residual demand curve, which causes it to bid more aggressively into the wholesale electricity market. This more aggressive bidding by the divested firm then presents all other suppliers with flatter residual demand curves, so they now find it optimal to submit flatter bid supply curves, which implies a flatter residual demand curve for the firm under consideration. Even in those cases when divestiture does not stop a supplier from being pivotal, the residual demand curve facing the firm that has less capacity should still be more elastic because more supply has been added to $SO(p)$, the aggregate bid supply function of all other firms besides the firm under consideration. This implies a smaller value for the firm's residual demand at all prices, as shown in Figure 3.

3.1.2. Forward Contracts and Vesting Contracts

Much has been made of the importance of forward contracts to manage the risk of spot price volatility. However, in electricity markets forward contracts serve an even more important role. They make it profit-maximizing for suppliers to bid more aggressively in the spot electricity market. This point is demonstrated in detail in Wolak (2000).

To understand the impact of forward contract commitments on supplier bidding behavior it is important to understand the obligations a forward contract for energy imposes on a supplier. Usually forward contracts are signed between suppliers and load-serving entities. These contracts typically give the load-serving entity the right to buy a fixed quantity of energy at a given location at a negotiated price. Viewed from this perspective, a forward contract for the supply of electricity obligates the seller to provide insurance against price volatility at a pre-specified location in the transmission network for a pre-specified quantity of energy. The seller of the forward contract does not have to produce energy from its own generating facilities to provide this price insurance to the purchaser of the forward contract. However, one way for the seller of the forward financial contract to avoid any price risk is to provide the contract quantity from its own generation units. This guarantees the firm will earn the difference between the forward contract price, *PC*, and its marginal cost, *MC*, times its contract quantity, *QC*, in variable profits (revenues in excess of variable costs) from the forward contract. This logic leads to another extremely important point about forward contracts that is not often fully understood by participants in a wholesale electricity market. Delivering electricity from a seller's own generation units is not always a profit-maximizing strategy given the supplier's forward contract obligations. This is also the reason why forward contracts provide strong incentives for suppliers to bid more aggressively (flatter bid supply functions) into the spot electricity market, which then leaves all other suppliers with more elastic residual demand curves.

To see this point, consider the following example taken from Wolak (2000). Let *DR(p)* equal the residual demand curve faced by the supplier with the forward contract obligation QC at a price of *PC* and a marginal cost of *MC*. For simplicity, I assume that the firm's marginal cost curve is constant, but this simplification does not impact any of the conclusions. The variable profits the firm earns during this hour are equal to

$$B(p) = (DR(p)-QC)(p - MC) + (PC - MC)QC. \tag{2}$$

The first term in (2) is equal to profit or loss the firm earns from buying or selling energy in the spot market at a price of p. The second term in (2) is the variable profits the firm earns from selling *QC* units of energy at *PC*. As discussed in Section 2, the firm's objective is to bid into the spot market in order to set a market price, *p*, that maximizes *B(p)*. Because forward contracts are, by definition, signed in advance of the operation of the spot market, from the perspective of bidding into the spot market, the firm treats *(PC - MC)QC* as a fixed payment it will receive regardless of the spot price, *p*. Consequently, the firm can only impact the first term through its bidding behavior in the spot market.

Because *DR(p)* is downward sloping, it is possible if the market price is high, the firm will sell less energy than its forward contract commitments. If the price at which *DR(p)* is less than *QC* is greater than *MC*, the firm incurs losses on the difference between *QC* and *DR(p)* times the difference between *p* and *MC*. Therefore, a supplier with a forward contract obligation of *QC* has a very strong incentive to submit bids that set prices below its marginal cost if it believes that *DR(p)* will be less than *QC*. This is because the supplier is effectively a net buyer of *QC - DR(p)* units of electricity, because it has already sold *QC* units in a forward contract. Consequently, it is profit-maximizing for the firm to want to purchase this net demand at the lowest possible price. It can either do this by producing the power from its own units at a cost of *MC* or purchasing the additional energy from the spot market. If the firm can push the market price below its marginal cost, then it is profit-maximizing for the firm to meet its forward obligations by purchasing power from the spot market rather than paying *MC* to produce it. Consequently, if suppliers have substantial forward contract obligations, they have extremely strong incentives to keep market prices very low until the level of energy they actually produce is greater than their forward contract quantity.

The competition-enhancing benefits of forward contract commitments from suppliers can be seen more easily by defining $DR_C(p) = DR(p) - QC$, the net-of-forward contract residual demand facing the firm and $F = (PC - MC)QC$, the variable profits from forward contract sales. In this notation, $B(p) = DR_C(p)(p - MC) + F$, which has exactly the same structure (except for *F*) as the firm's profits from selling electricity if it had no forward contract commitments. The only difference is that *DR(p)* replaces $DR_C(p)$ in the expression for the supplier's variable profits. Consequently, profit-maximizing behavior implies that the firm will submit bids to set a price in the spot market that satisfies equation (1) with *DR(p)* replaced by $DR_C(p)$. This implies the following relationship between P^c, the expected profit-maximizing price, the firm's marginal cost of production, *MC*, and $,^c$, the elasticity of the net-of- forward-contract-quantity residual demand curve evaluated at P^c:

$$(P^c - MC)/P^c = - 1/,^c, \tag{4}$$

where $,^c = DR_CN(P^c)*(P^c/DR_C(P^c))$. Because $DR_C(p) = DR(p) - QC$, this implies that at the same market price, *p*, and residual demand curve, *DR(p)*, the absolute value of the elasticity of the net-of-forward-contract-quantity residual demand curve is always greater than the absolute value of the elasticity of the residual demand curve. Simple proof of this result follows from the fact that

$DR_CN(p) = DRN(p)$ for all prices and $QC > 0$, so that by re-writing the expressions for $,^c$ and „ we obtain:

$$| ,^c | = | DRN(p)*(p/[DR(p) - QC]) | > | , | = | DRN(p)*(p/DR(p)) |. \qquad (5)$$

Moreover, as long as $DR(p) - QC > 0$, the larger the value of QC, the greater is the difference between $,^c$ and „ and the smaller is the expected profit-maximizing percentage mark-up of the market price above the firm's marginal cost of producing the last unit of electricity that it supplies with forward contract commitments versus no forward contract commitments. This result demonstrates that it is always unilateral profit-maximizing, for the same underlying residual demand curve, for the supplier to set a lower price relative to its marginal cost if it has forward contract commitments.

This incentive to bid more aggressively in the spot market if a supplier has substantial forward contracts also has implications for how a fixed quantity of forward contract commitments should be allocated among suppliers to maximize the benefits of these contracts to the competitiveness of the spot market. Because a firm with forward contract obligations will bid more aggressively in the spot market, this implies that all of its competitors will also face more elastic residual demand curves and therefore find it unilaterally profit-maximizing to bid more aggressively in the spot market. This more aggressive bidding will leave all other firms with more elastic residual demand curves, which should therefore make these firms bid more aggressively in the spot market.

This virtuous cycle with respect to the benefits of forward contracting implies that a given amount of forward contracts will have the greatest competitive benefits if it is spread out among all of the suppliers in the market roughly proportional to their generation capacity ownership shares. For example, if there are five firms and each of them owns 1,000 MW of capacity, then forward contract commitments should be allocated equally across the firms to maximize the competitive benefits. If one firm owned twice the capacity of other firms, then it should have roughly twice the forward contract commitments to load-serving entities that the other suppliers have.

Because of the spot market benefits of substantial amounts of forward contract commitments between suppliers and load-serving entities, most wholesale electricity markets begin operation with a large fraction of the final demand covered under forward contracts. If a substantial amount of capacity is initially controlled by government-owned or privately-owned monopolies, the regulator or market designer usually orders that most of these assets be sold to new entrants to create a more competitive wholesale market. These sales typically take place with forward contract commitments on the part of the new owner of the generation capacity to supply a substantial fraction of the expected output of the unit to load-serving entities at some pre-set price. These contracts are typically called vesting contracts, because they are assigned to the unit as a pre-condition for its sale. For example, if a 500 MW unit owned by the former monopolist were being sold, the regulator would assign a forward contract obligation on the new owner to supply 400 MW of energy each hour at some previously agreed upon price to one of the load-serving entities.

Vesting contracts accomplish several goals. The first is to provide price certainty for load-serving entities for a significant fraction of their wholesale energy needs. The second is to provide revenue certainty to the new owner of the generating facility. With such a forward contract, the new owner of the generation unit in our example already has a revenue stream each hour equal to the contract price times 400 MWh. These two aspects of vesting contracts protect suppliers and loads from the vagaries of spot market outcomes, because both parties receive or pay the spot price for production or consumption beyond the contract quantity. Finally, the existence of this forward contract obligation has the beneficial impacts on the competitiveness of the spot energy market described above.

The major cause of the California electricity crisis is the fact that California's three large load-serving entities purchased 100 percent of their total energy and ancillary service requirements from the day-ahead and shorter-horizon spot markets. When the amount of imports from the Pacific Northwest was substantially reduced as a result of lower water availability during the late spring and summer of 2000, the fossil fuel suppliers found themselves facing the significantly less elastic residual demand curves for their output documented in Wolak (2003b). This fact made the unilateral profit-maximizing mark-up of price above the marginal cost of producing electricity substantially higher during the summer and autumn of 2000 than it had been during the previous two years of the market. Moreover, particularly during the latter part of the autumn of 2000, there is strong evidence that the marginal cost of highest cost unit operating in California increased substantially relative to the early part of 2000 and the previous two years.

During the vast majority of hours of the year, natural gas-fired units set the price in California. During the latter part of 2000, the price of natural gas increased substantially relative to the levels that existed in early 2000 and the previous two years. These substantially higher natural gas prices resulted in much higher values for the marginal cost of the highest cost unit operating in California—market price under the assumption that all suppliers were unable to exercise any unilateral market power. Applying hourly values of the residual demand facing the five large suppliers in the latter parts of 1999 and 1998 to these higher marginal costs would have led to substantially higher prices in 2000. The combination of these higher natural gas prices and the substantially smaller average hourly values of elasticity of the residual demand curve facing each of the five large suppliers in the California electricity market in the latter part of 2000, calculated in Wolak (2003b), led to even higher prices during the latter part of 2000.

Wolak (2003b) provides a detailed diagnosis of causes and consequences of the California electricity crisis and concludes with two very important lessons that are very relevant to LACs. The first is the importance of an active forward market for electricity. Although vesting contracts are one way to foster an active forward market during the initial stages of a restructuring process, they are not essential to success of a restructuring process. What is essential, particularly for a market that normally obtains a substantial fraction of its electricity from hydroelectricity, is that load-serving entities purchase a substantial fraction of their energy needs, certainly more than 80 percent, in the forward market at least a year in advance of delivery. The second lesson is the necessity of a credible regulatory process that sets well-defined boundaries on acceptable market outcomes, defines those outcomes that justify regulatory invention, clarifies in advance the form of this regulatory intervention, and follows through with these pre-commitments should the standards for acceptable market outcomes be violated.

3.1.3. Involving Final Demand

If there were no variation in demand or supply across hours of the day, days of the week or weeks of the year, it would be possible to build enough generation capacity to ensure that all demand in every hour could be served at some fixed price. However, the reality of electricity consumption and generation units and transmission network operations is that demand and supply vary over time, often in an unpredictable manner. This implies that there is always some likelihood that available capacity will be insufficient to meet demand.

Given available generation capacity, there are two ways of eliminating this imbalance: either price must be increased so as to choke off demand, or demand must be rationed. Rationing is clearly an extremely inefficient way to ensure that supply equals demand. Many consumers willing to purchase electricity at the prevailing price are unable to do so. Moreover, as has been discovered by politicians in all countries where rationing has occurred, the backlash associated with rationing can be devastating to those in power. Moreover, the indirect costs of rationing on the level of economic activity can be substantial. In particular, preparing for and dealing with rationing periods causes substantial losses in economic activity.

A far superior approach to dealing with a shortfall of available supply relative to the level of demand at the prevailing price is to allow this retail price to rise to the level necessary to cause a sufficient number of consumers to reduce their consumption so that supply and demand balance. Although this might seem like a revolutionary concept in the electricity supply industry, this is precisely how markets for all other products operate.

This concept is even less revolutionary when one recognizes that from a system reliability perspective, customers paying the hourly price of wholesale electricity for their hourly consumption is not fundamentally different from generation unit owners being paid according to the hourly price of electricity for their hourly production. Let $D(p)$ equal the consumer's hourly demand for electricity as a function of the hourly price of electricity. Define $SN(p) = D(0) - D(p)$, where $D(0)$ is the consumer's demand for electricity at an hourly price equal to zero. The function $SN(p)$ is the consumer's true willingness supply curve for "negawatts." Because $D(p)$ is a downward-sloping function of p, $SN(p)$ is an upward-sloping function of p. A generator with a marginal cost curve equal to $SN(p)$ has the ability to provide the same reliability benefits to this consumer. However, as discussed above, an electricity supplier has the incentive to maximize the profits it earns from selling electricity in the spot market given its marginal cost function. In contrast, I would expect an industrial or commercial consumer with a true supply curve of negawatts, $SN(p)$, to bid her willingness to supply negawatts into the spot market to maximize the profits associated with selling her final output, which would imply demand-bidding to reduce the market price. Consequently, even though the generator and consumer have the same true willingness to supply megawatt and negawatts, respectively, each of them will use this true supply curve in a different manner. The supplier will use it to exercise market power on the supply side to raise market prices, and the consumer will use it to exercise market power on the demand side of the market to reduce the price it pays for electricity. Wolak (2001) describes how a load-serving entity with some consumers facing the hourly wholesale price or a large consumer facing the hourly price could exercise market power on the demand side to reduce the average price it pays for a fixed daily supply of electricity consumed over longer period of time.

Besides allowing the system operator more flexibility in managing demand and supply imbalances, the presence of some consumers that alter their consumption in response to the hourly wholesale price also significantly benefits the competitiveness of the spot market. Figure 4 illustrates this point. The two residual demand curves are computed for the same value of $SO(p)$. One, Q_D, is perfectly inelastic. The other, $Q_D(p)$, is price elastic. As shown in the diagram, the slope of the resulting residual demand curve using $Q_D(p)$ is always flatter than the slope of the residual demand curve using Q_D. Following the logic used for the case of forward contracts, it can be demonstrated that for the same price and same value of residual demand, the elasticity of the residual demand curve using $Q_D(p)$, is always greater than the one using Q_D, because the slope of the one using $Q_D(p)$ is equal to $DRN(p) = Q_DN(p) - SON(p)$, which is larger in absolute value than $SON(p)$, the slope of the residual demand curve using Q_D. Consequently, the competition benefit of having final consumers pay the hourly wholesale price is that all suppliers will face more elastic residual demand curves, which will cause them all to bid more aggressively into the spot market.

Politicians and policymakers often express concern that subjecting consumers to real-time price risk will introduce too much volatility into their monthly bill. These concerns are, for the most part, unfounded as well as misplaced. Some entity must manage wholesale spot price risk. Just because a regulator sets a fixed price or pattern of prices throughout the day (time-of-use prices), some entity must still ensure that over the course of the month or year, the retailer's total revenues less transmission, distribution and supply costs, must cover total wholesale energy costs. If the regulator sets this fixed price too low relative to the current wholesale price, then either the retailer or the government must pay the difference. If this revenues shortfall relative to costs continues, eventually the government must make up the difference because it has the ability to impose taxes to fund its expenditures. However, these tax revenues are also collected from consumers of electricity, although not generally in proportion to their consumption of electricity.

This is the lesson learned by the citizens of California during the period from May 2000 to June 2001. When average wholesale prices rose above the average wholesale price implicit in the frozen retail price California consumers paid for electricity, retailers initially made up the difference. Eventually, these companies threatened to declare bankruptcy in the case of Southern California Edison and San Diego Gas and Electric, and in fact declared bankruptcy in the case of Pacific Gas and Electric, so that the state of California had to take over purchasing wholesale power at even higher prices. The California experience demonstrates that the option to purchase all retail electricity demand at a price that does not vary with hourly system conditions is extremely valuable to consumers, and if exercised it can be extremely costly to the government that must ultimately finance it.

This is a restatement of a standard prediction from the theory of stock option valuation: the price of a call option on a stock increases with the volatility of the underlying security. However, the fact that consumers have the option to switch to this fixed-price, full-requirements rate that completely shields them from any spot price risk in their electricity purchases (but not in their tax payments) makes wholesale prices more volatile. This follows from the fact that a less elastic wholesale electricity demand increases price volatility. Clearly, a more efficient way to manage electricity spot price risk is to treat consumers the same way that generation unit owners are treated because, as discussed above, consumers paying hourly prices have the potential to provide the same level of grid reliability benefits as generation unit owners.

Perhaps the most important, but most often ignored, lesson from electricity restructuring processes in developed countries is the necessity of treating load and generation symmetrically. Unfortunately, very few developed countries do this, which may explain why few of them have seen consumers realize significant benefits from ESI restructuring. Symmetric treatment of load and generation means that unless a retail consumer signs a forward contract with an electricity retailer the default wholesale price the customer pays for all of his consumption is the hourly wholesale price. This is precisely the same wholesale price risk that generation unit owners face. Unless they sign a forward contract with a load-serving entity or some other market participant, the price they receive for any real-time energy production is the hourly spot price. Just as very few suppliers are willing to risk selling all of their output in the spot market, I would expect consumers to have similar preferences against too much reliance on the spot market and would therefore be willing to sign a long-term contract for a large fraction of their expected hourly consumption during each hour of the month. For example, a residential consumer might purchase a pre-specified daily or weekly energy profile at a fixed price for the next 12 months. This customer would then be able to sell energy she has purchased but does not consume during any hour at the hourly wholesale price or purchase any power she needs beyond the contract quantity at that same price. This type of pricing arrangement would result in a significantly less volatile monthly electricity bill than if the consumer made all of his purchases at the hourly wholesale price.

If all customers purchased according to this sort of pricing plan, then there would be no residual spot price risk for the government to manage using tax revenues. Instead, all consumers manage the risk of high wholesale prices, according to their preferences for taking on spot price risk. Those willing to take on more spot price risk could do by purchasing a smaller fraction of their expected hourly demand as a fixed-price contract. Moreover, because all consumers have an incentive to reduce their consumption during high-priced periods and shift it to lower-priced periods, wholesale prices are likely to be significantly less volatile. Rather than continuing to consume when wholesale prices rise, they now see this very high spot price as the opportunity cost of using electricity for all of their consumption, with the important difference that if they consume less than their forward contract quantity, they are paid this very high price for each KWh they do not consume below that level.

Symmetric treatment of load and generation does not preclude a customer from purchasing a fixed-price full requirements contract for all of the electricity he might consume in a month. It only imposes the requirement that the consumer pay the full cost of supplying this product. The wholesale energy cost portion of such a contract should be substantially higher than the expected average wholesale price over the duration of the contract because the supplier of the contract is providing insurance against fluctuations in the wholesale price for a quantity of energy that is determined at the discretion of the customer, meaning that there is both price and quantity risk associated with providing a full-requirement, fixed-price contract. To get an idea of the risk involved, imagine a gasoline retailer making a promise to its customers that they can purchase as much gasoline as they would like at a fixed price for an entire year. Given the volatility in wholesale gasoline prices, the premium (above the expected average price) that a retailer would require to offer such a product should be quite high. This same sort of price premium should also exist for full requirements, fixed-price contracts for electricity.

Having a regulatory process determine the fixed-price, full-requirements rate runs into the following problems. If the regulatory process sets this price too low, then too many customers will choose this rate and the retailer will be unable to procure all of the wholesale energy necessary to supply these contracts and at wholesale prices that allow full cost recovery. In the extreme case this is the California problem described above. However, a number of markets in the eastern United States have experienced less extreme versions of this problem. Alternatively, if the regulatory process sets this price too high, then too few customers will select this supply arrangement. The cost to customers of setting this rate too high may be very low or even zero, if retailers are allowed to compete in offering fixed-price, full-requirements supply contracts and no customer ultimate elects to take supply at this regulated rate. Consequently, one solution to the problem of providing sufficient spot price protection for customers in an environment with retail competition is for the regulator to determine periodically a slack upper bound on the rate for a fixed-price, full-requirements contract of a fixed duration and make it a pre-condition to be an electricity retailer to offer such a contract along with all of the other supply contracts the retailer would like to offer.

Borenstein (2003) discusses a number of technical issues associated with involving final demand in the retail market. One roadblock to symmetric treatment of load and generation for all electricity customers that Borenstein treats is the cost of installing the necessary metering technology at the household level to allow consumption to be measured on an hourly versus monthly basis. Wolak (2001) presents evidence for California that suggests that these costs could be paid for by the lower wholesale electricity prices that would result from the more competitive wholesale market brought about by the symmetric treatment of load and generation. Green and McDaniel (1998) perform a social cost-benefit analysis of the transition to retail competition for residential consumers in the England and Wales electricity market where any consumer that wishes to switch from their default supplier must install a half-hourly meter. Green and McDaniel analyzed a number of likely scenarios for the impact of retail competition on residential consumers and electricity suppliers and found that the net benefits, if any, are likely to come later as more consumers participate in the retail market and competitive pressures reduce retail prices. Nevertheless, all of these researchers argue that there are significant benefits net of metering costs from involving commercial and industrial consumers in the wholesale market.

3.1.4. Economic Reliability versus Engineering Reliability of a Transmission Network

The presence of a wholesale electricity market changes the definition of what constitutes a reliable transmission network. As shown in Section 2, in order for it to be profit-maximizing for a generation unit owner to submit a bid supply curve close to its marginal cost curve, the supplier must face a sufficiently elastic residual demand curve for its energy. For this to be the case, there must be enough transmission capacity into the geographic area served by this unit owner so that any attempts it makes to raise local prices will result in enough lost sales to make such a bidding strategy unprofitable.

I introduce the concept of an economically reliable transmission network as one with sufficient capacity so that each location in the network faces sufficient competition from distant generation to cause the local unit owners to compete with distant generators rather than withhold energy to cause congestion into the region and create a local monopoly market. In the former vertically integrated monopoly regime, transmission expansions were undertaken to ensure the engineering reliability of the transmission network. A transmission network was deemed to be reliable from an engineering perspective if the vertically integrated monopolist that controlled all of the generation units in the control area could maintain a reliable electricity supply to consumers despite unexpected generation and transmission outages.

The value of increasing the transmission capacity between two points still depends on the extent to which this expansion allows the substitution of cheap generation in one area for expensive generation in the other area. Under the vertically integrated monopoly regime, all differences across regions in wholesale energy charges were due to differences in the local costs of production for the geographic monopolist. However, in the wholesale market regime, the extent of market power that can be exercised by firms at each location in the network can lead to much larger differences in payments for wholesale electricity across regions. For example, even if the difference in the variable cost of the highest cost units operating in two regions is less than $10/MWh, because firms in one area are able to exercise local market power, differences in the wholesale prices that consumers must pay across the two regions can be as high as the price cap on the real-time price of energy. For example, during early 2000 in the California market when the price cap on the ISO's real-time market was $750/MWh, because of congestion between the SP15 and NP15 zones, prices in the two zones differed by as much as $700/MWh. This occurred in despite the fact that the difference in the variable costs of the highest cost units operating in the two zones was less than $10/MWh.

This example demonstrates that a major benefit of transmission capacity in a wholesale market regime is that it limits the ability of generation unit owners to use transmission congestion to limit the number of competitors they face. More transmission capacity into a local area implies that local generating unit owners face more competition from distant generation for a larger fraction of their capacity. Because these firms now face more competition from distant generation, they must bid more aggressively (lower prices) over a wider range of local demand realizations to sell the same amount of energy as they did before the transmission upgrade. In all cases, this more aggressive bidding brought about by the transmission upgrade will lower average wholesale energy prices on the congested side of the interface. Moreover, to the extent that the probability of congestion in one direction on an interface is approximately equal to the probability of congestion in the opposite direction, the reduced opportunities for suppliers to exercise market power on both sides of the interface as a result of a transmission upgrade could reduce average wholesale prices at both locations.

The opportunity for generation unit owners to impact location prices through their scheduling and bidding behavior creates another source of benefits of transmission upgrades in the wholesale market regime. In the vertically integrated monopoly regime, one rationale for upgrades of the monopolist's network was to manage the reliability risk associated with generation or transmission line outages. For example, an upgrade could be justified by the logic that, if certain generating units became unavailable, the supply shortfall could be temporarily served with distant, but more expensive, generating units. The reliability justification for such upgrades was that the cost of upgrading was less than the economic value created by the additional electricity that the consumers were able to consume because of the transmission upgrade.

Under the competitive market regime generators may have an additional incentive, besides the fact that the unit is physically unable to operate, to declare their unit unavailable. They may find it profitable to create an artificial scarcity of generating capacity in a geographic area in order to increase the wholesale price they receive for the energy they do sell. This incentive to withhold generating capacity did not exist in the regulated monopoly regime. The monopolist was required by law to serve all demand at the regulated retail price. Declaring a unit unable to operate would only reduce the monopolist's expected profits, because it has the same load obligations, but now it has fewer units to choose from to serve these obligations. However, in the wholesale market regime, if a generator is able to raise the price it receives for its power by 100 percent by withholding (declaring unavailable to operate) a small fraction of its capacity, it will find this behavior profitable.

Consequently, in the wholesale market regime, reliability risk has an additional dimension because of the incentive for generation unit owners to withhold capacity from the market to increase prices if they do not face sufficient competition. For example, few, if any, market observers would have predicted as late as August 2000 that the California ISO would experience a daily average of approximately 10,000 MW of generating units off-line during the eight-month period November, 2000 to May, 2001. However, as discussed in Wolak (2003c), these outage levels made it considerably more straightforward for suppliers to raise the prices they received for the energy they actually produced. Additional transmission capacity can render physical withholding strategies, which may lead to load curtailments, less profitable and therefore less likely to occur.

Understanding how transmission upgrades can increase the elasticity of the residual demand curve a supplier faces requires only a slight modification of the discussion surrounding Figure 3. Suppose that 9,500 MWh of demand is all located on the other side of a transmission line with 9,000 MW of capacity, and the supplier under consideration owns 1000 MW of generation local to the demand. Suppose there are 12 firms, each of which owns 1,000 MW of capacity located on the other side of the interface. In this case, the local supplier is pivotal for 500 MWh of energy because local demand is 9,500 MWh, but only 9,000 MWh of energy can get into the local area because of transmission constraints. The assumption of 12 firms, each of which owns 1,000 MW of capacity, implies that there is plenty of generation available to serve the local demand. It just cannot get into the region because of transmission constraints. This fact allows a re-interpretation of $SO_1(p)$ in Figure 3 as the bid supply curve of suppliers to sell energy along the 9,000 MW of transmission capacity.

Suppose the transmission line is now upgraded to 9,500 MW. From the perspective of the local firm this results in a $SO_2(p)$ to serve the local demand that has an additional 500 MW, which means that the local supplier is no longer pivotal. Before the upgrade the local supplier faced the residual demand curve $DR_1(p)$ in Figure 3 and after the upgrade it faces $DR_2(p)$, which is more elastic than $DR_1(p)$ at all price levels. This is the mechanism by which transmission upgrades increase the competitiveness of wholesale electricity markets.

3.1.5. Specific Market Design Lessons from Developed Countries

The three best-performing developed country wholesale electricity markets—Australia, England and Wales, and the Nordic countries—have achieved competitive wholesale markets by implementing one or more of the prescriptions described above for increasing the elasticity of the residual demand curves that suppliers face.

For example, since the start of the England and Wales market, the regulator has ordered a series of capacity divestitures from National Power and PowerGen, the two large fossil-fuel generation companies established at the start of the market. In the State of Victoria in Australia, the generation capacity of the former monopoly supplier was sold at the plant level to ensure a sufficient number of competing suppliers. There has been very little divestiture of the generation capacity owned by the large government-owned firms in the Nordic markets. However, as discussed in Section 2, it is unclear whether these firms have as strong of an incentive to maximize profits as privately owned firms, which could explain why their dominant position in this market has not led to the exercise of excessive levels of unilateral market power.

In all three of these markets the vast majority of final demand is covered by long-term contracts. It is difficult to get precise estimates of the extent of final demand that is covered by long-term contracts, because this information is considered proprietary. However, 85 percent is a conservative lower boundary on the percent of final demand covered by long-term contracts for all of these markets. Green (1999) presents estimates of these magnitudes for the mid-1990s for the England and Wales electricity market. Wolak (2000) discusses the situation in the Australian electricity market in the mid-1990s. Wolak (1999) describes conditions in the Nordic market during the mid-1990s. All of these authors find forward contract coverage percentages above this level.

The development of an active demand side has been much slower in all of these countries. The England and Wales market and the Nordic market are the only countries that treat all load and generation symmetrically. Australia is still in the process of moving forward with retail competition for all consumers.

Finally, all three countries have transmission networks that experience very little transmission congestion. Although these transmission networks were not built to serve a competitive wholesale market, they are certainly far better suited to this task than the transmission networks that exist in the United States and many other developed countries that have restructured. In England and Wales, National Grid in England and Wales is a privately-owned firm that owns and operates the transmission network and the balancing mechanism for the England and Wales electricity market. Since restructuring took place in the early 1990s, National Grid has undertaken a number of upgrades to its transmission network. Because all of the states in Australia had their own government-owned monopoly ESIs before they restructured, the transmission networks in each of these states were designed to facilitate the delivery of electricity anywhere in the state. Consequently, the Australian market only experiences significant congestion across state boundaries. The Nordic market has a similar history. Most of the Nordic countries built extensive transmission networks within their boundaries, so that congestion primarily occurs across national boundaries.

The fact that transmission congestion in these markets tends to occur primarily across national or state boundaries implies that for most hours of the year there is sufficient competition among suppliers within these geographic regions so that no single supplier is able to exercise significant unilateral market power. Unfortunately, this is not the case in a number of LACs. The significantly higher frequency of congestion within most LAC countries gives rise to a problem that is increasingly relevant to the United States: the local market power problem. The United States experience has shown that until more transmission capacity or generation capacity can be built, regulatory intervention is the only solution to the local market power problem.

3.1.6. Credible and Effective Regulatory Process

Any attempt to establish a competitive market without the conditions outlined in the previous four sub-sections is bound to result in periods when the market fails in unintended ways. This is particularly true in a market with significant local market power problems, as is likely to be the case in most LACs, because the country's existing transmission network was not designed to achieve the economic reliability standard of a wholesale market regime. For this reason, it is essential that there be a credible and effective regulatory process in place to mitigate substantial local market power and to monitor overall market performance to detect and correct market design flaws while they only cause limited consumer harm. Unlike the case of the vertically integrated monopoly regime, the regulator must be forward-looking and fast-acting because, as emphasized in Section 2, markets provide extremely high-powered incentives for firm behavior, so it does not take very long for a wholesale electricity market to cause enormous consumer harm. The California electricity crisis is an example of this phenomenon. The Federal Energy Regulatory Commission (FERC), the entity that regulates wholesale markets in the US, waited almost six months from the time it first became clear that there was substantial market power being exercised in the California market before it took action. In addition, the regulatory intervention in December 2000 was so timid and ill-conceived that its result was to increase the rate at which consumer harm occurred. Wolak, Nordhaus, and Shapiro (2000) discuss the likely impact, which also turned out to be the eventual impact, of the FERC's December 2000 action.

An argument, based on the logic of the individual rationality constraint in Section 2, can even be made that an effective, credible and fast-acting regulatory process will increase the competitiveness of a wholesale electricity market. Specifically, if the regulator makes the penalties associated with any market rule violations more than the benefits that the market participant receives from violating that market rule, then suppliers will find it profit-maximizing to obey the market rules. One lesson from the activities of many firms in the California market and other markets in the US is that if the cost of a market rule violation is less than the benefit the firm receives from violating the market rule, the firm will violate the market rule and pay the associated penalties as a cost of doing business because it earns a positive profit as a result of these actions.

FERC's failure to recognize this logic allowed the California electricity crisis to last as long as it did and become as big of a disaster as it did. Since the start of the California market FERC refused to implement a system of financial penalties for market rule violations. FERC only required firms to pay back the so-called ill-gotten gains from market rule violations. Clearly, this approach does not deter profitable market rule violations because the worst-case scenario for the firm is giving the profits back and the best case is being able to keep them. Unless the regulator is flawless at detecting market rule violations, under these circumstances it is expected profit-maximizing for the firm to violate market rules because it earns zero if it is caught violating the rules and positive profits when it does violate the market rules. This is not the incentive for firm behavior a regulator wants to create. Unfortunately, this is precisely the incentive that FERC created in the California market.

Any penalty mechanism the regulator implements should accomplish two goals. First, firms should pay fines for market rule violations that at least exceed the financial damages its actions impose on other market participants. Second, this penalty should also be sufficient to make the expected amount of fines the firm must pay as a result of violating a market rule greater than the expected benefit the firm obtains from this violation. This second constraint implies that the firm finds it profit-maximizing to comply with the market rules. A regulator that does not take decisive action to penalize market rule violations subject to these two constraints on the magnitude of fines imposed will soon find market rule violation more frequent, which will make it more costly for the ISO to manage the transmission network and operate its energy and ancillary services markets.

The experience of California and all other US states with wholesale markets provides another very valuable lesson for the design of an effective and credible regulatory process. Retail market regulatory policies must be consistent with wholesale market regulatory policies or wholesale market outcomes that are harmful to consumers and ultimately producers will occur. In the US this is a particularly challenging task because of the division of regulatory responsibilities: federal regulators are responsible for wholesale markets and state regulators are responsible for retail markets. FERC regulates wholesale electricity markets throughout the US. It also has far more ambitious plans for wholesale electricity markets than any of the state public utilities commissions (PUCs) that regulate retail electricity markets within their boundaries. As a result, the state PUCs often enact retail market policies that work against many of the necessary features of a competitive wholesale market described above. For example, they do not allow load and generation to be treated symmetrically. In fact, they usually do all they can to shield consumers from electricity prices that vary over time or location. In addition, state PUCs also tend to be skeptical of transmission upgrades to facilitate a more competitive wholesale market. Finally, they are often reluctant to give the load-serving entities they regulate sufficient freedom to engage in the necessary forward financial contracts to manage their wholesale market spot price risk effectively. Pairing these retail market policies with FERC's very progressive wholesale market policies, which appear to assume that load-serving entities have substantial flexibility to forward contract and that final consumers are active participants in the wholesale market, can create disasters like California and the smaller, but still very costly, market failures that have occurred in the PJM, New York and New England ISOs.

However, the requirement to coordinate wholesale and retail market policies has a very important implication that should guide the reform process in both developed and developing countries. If a country is not going to follow the five recommendations for a competitive wholesale market outlined in Section 3.1, then it must have substantially less ambitious goals for its wholesale electricity market.

For instance, if the political process is unwilling to divest enough of the capacity of the largest supplier to new entrants, this should place limits on the form and operation of the wholesale market. If the regulator or political process is unwilling to allow retailers sufficient flexibility to manage their spot price risk or to require some or all final consumers to be treated symmetrically with generation unit owners in the wholesale market, this should constrain the type of wholesale market adopted. These constraints on the wholesale market should not be relaxed until the regulatory constraints on achieving the five goals outlined above are relaxed. Similar logic applies to a country or region that refuses to consider the economic reliability benefits of transmission upgrades in the cost-benefit calculus for transmission upgrades. Finally, a country that is unwilling to establish an independent regulator or regulatory body with the necessary statutory powers to become credible and effective should not even consider reforming its ESIs.

The regulatory body is the guiding force for the reform process. Unless the regulator is able to implement a local market power mitigation mechanism and to intervene and change harmful market rules or market structures, subject to judicial review, significant consumer harm is likely to occur at some point in the future. Establishing a credible and effective regulatory process in a developing country with no history of regulation is perhaps the most technically and politically challenging task in establishing competitive wholesale electricity markets in LACs.

3.2. Implementing Workably Competitive Markets in Developed Countries

Although the previous section provides a roadmap for designing a wholesale electricity market that is as competitive as possible given the set of available technologies for producing, transmitting and distributing electricity, all developed countries that have restructured their ESIs have introduced wholesale competition without one or all of the necessary features described in Section 3.1. This has created a host of market failures, the most notable being the California electricity crisis. In addition, the substantial market power exercised in the England and Wales electricity market throughout the 1990s was the result of an attempt to introduce wholesale competition without these necessary features. The initial market design failure in the England and Wales market was due to insufficient competition among suppliers and the market rules for dispatching and paying generation unit owners that were taken directly from the former monopoly regime, without sufficient consideration of their impact on supplier behavior in the wholesale market regime. Patrick and Wolak (1997) discuss one such market rule and its impact on the ability of suppliers to exercise market power in the England and Wales electricity pool.

No country has been able to avoid market failures of some magnitude as a result of implementing a wholesale market without the necessary initial conditions described above. Therefore, it seems more realistic to expect that countries will continue to introduce wholesale markets with only a few, if any, of these necessary features in place. This also seems likely for developing countries as well. Consequently, the goal of this section is to describe a number of safeguards that countries have used to guard against adverse wholesale market outcomes.

Because it is impossible to characterize all possible safeguards against less-than-optimal initial conditions in a wholesale electricity market, I will focus on the intersection of safeguards that have been most commonly implemented and those most relevant to the experiences of developing countries. This discussion will describe how these safeguards have allowed the wholesale market to achieve tolerable outcomes while improvements in the market design along the five dimensions described in Section 3.1 can be implemented. I will also point out the trade-off built into all of these measures. Typically, they protect consumers from harmful market outcomes in the short term by limiting the potential long-term benefits to consumers from a wholesale electricity market.

3.2.1. The Role of Government-Owned Market Participants

Significant participation of government-owned entities in the wholesale market, particularly in the generation sector, can limit the potential for consumer harm. As discussed in Section 2, the incentive for managers of these firms to maximize profits is significantly more muted than that facing the managers of privately owned firms.

One explanation for the lack of a significant market failure in the Nordic market is the dominant market position of two government-owned companies—Statkraft in Norway and Vattenfall in Sweden. In Australia, all of the New South Wales generation units are owned by the state. The Snowy Mountains hydro project on the border of New South Wales and Victoria is jointly owned and operated by these two states. Substantial government participation in the market provides more avenues for the regulator and political process to intervene in the market and limit the damage to consumers caused by market design flaws. For example, the political process can impose restrictions on the behavior of government-owned entities that it is unable to impose on privately owned companies.

Clearly, there is a downside to this safeguard. The experience of Norway is instructive on this issue. Wolak (1999) discusses an instance early in the Nordic market when Statkraft announced that wholesale electricity prices were too low and that it would not release any water until prices rose, which they subsequently did. However, the opposite circumstance could have occurred if Statkraft felt political pressure to reduce electricity prices. This would cause Statkraft to release water to drive prices down. Privately owned potential entrants to the Nordic market recognize this dominant position of Statkraft and its implications for the profitability of fossil-fuel-based entry. As a result, there have been no significant new generation capacity investments in Norway since the wholesale market was started in the early 1990s. The chilling impact on private investment in new generation capacity in markets dominated by the state-owned firms is particularly relevant in hydro-dominated systems where the variable cost of producing electricity is zero for most hours, so the only cost of selling water as electricity is the foregone profit from selling it in another hour.

Although significant government participation in the generation sector may protect consumers from substantial harm, it has a cost in terms of long-term market efficiency. For the reasons discussed in Section 2, I would not expect government-owned suppliers to produce their output in a minimum-cost manner or undertake the new investments necessary to meet demand growth in a minimum-cost manner. The evidence from the developed country reforms is that fossil-fuel based markets with few, if any, government-owned firms have experienced the largest amount of private investment in new generating facilities. In a particularly pronounced example, in the England and Wales market, a substantial increase in new combined cycle gas turbine capacity has driven wholesale prices so low that many generation unit owners claim to be on the brink of filing for bankruptcy protection. The US has seen a similar pattern. A substantial amount of new investment in generation capacity occurred in the parts of the US with formal wholesale markets. Even the California market has had significant new capacity investments, with roughly a 10 percent increase in generation capacity on line over the past two years.

The combination of new investments and slower growth in demand because of a slowdown in the US economy has led to very low wholesale electricity prices throughout the US, which has adversely impacted the financial health of US wholesale electricity suppliers. In contrast, Victoria and New South Wales, the most populous states in Australia, have seen no new investment, except in peak period generation facilities, despite starting the decade of the 1990s with close to the same reserve margin as England and Wales.

Consequently, one strategy for reform is to allow state ownership initially and then sell off assets as many of the kinks are worked out in the initial market design. Implementing market rule changes in a market dominated by state-owned entities tends to be much easier, because most market rule changes reallocate revenues across firms. If most of the revenues go to government-owned firms, the distributional issues associated with implementing market rule changes should be less contentious. If a country decides to take this path, it is important to recognize the difficulty in selling off government-owned assets once the politicians and regulators realize that intervention in the market will be more difficult and less effective with less government-owned firms. Consequently, this short-term safeguard could easily turn into a permanent barrier to competition in the long-term.

This path may be particularly problematic for LACs interested in forming a wholesale market to attract private investment. The combination of a generation fleet dominated by hydroelectric capacity and a dominant ownership position by the government may be counterproductive to the goal of the reform process. It is reasonable to expect private ESI investors in LACs to have the same fears about the behavior of the government-owned hydroelectric capacity as potential investors in the Norwegian market. My recommended market design for LACs will provide suggestions for addressing the conflicting goals of protecting consumers and encouraging private investment in new capacity.

3.2.2. Cost-Based Dispatch and Local Market Power Mitigation

A number of markets with transmission networks poorly suited for wholesale competition have addressed this problem by implementing cost-based local market power mitigation mechanisms or simply cost-based dispatch schemes. The local market power mitigation mechanisms usually give the ISO the discretion to determine, depending on current system conditions, whether a supplier possesses substantial local market power. Under those circumstances, the bids of units deemed to possess local market power are mitigated to their variable cost, or variable cost plus an adder, and these mitigated bids are then entered into the ISO's price-setting process. In a market with cost-based dispatch, suppliers are required to file their start-up, no-load and variable costs with the regulator. Once these costs are approved by the regulator, suppliers are required to bid these costs into the market, and they are paid market-clearing prices based on these cost-based bids.

The logic behind these mitigation mechanisms follows from the discussion of impact of transmission network expansion on the elasticity of the residual demand curve facing a supplier from Section 3.1.4. Specifically, the ISO must judge whether there exists sufficient competition among suppliers on either side of a congested interface to allow market forces to set the price on both sides of the interface. Viewed from this perspective, transmission constraints shrink the number of suppliers that compete to serve demand at each location in the network. Almost by definition, a higher incidence of transmission congestion cannot lead to more competitive bidding by suppliers.

To understand how a commonly used local market power mitigation mechanism works, consider the example in Section 3.1.4 with 9,000 MW of transmission capacity into the demand center with 9,500 MWh of local demand and 1,000 MW of local generation. Under these circumstances, the local supplier is pivotal for 500 MWh of energy. Allowing this local monopolist to set the market price will cause it to submit an extremely high bid. If there is no price cap in this market, there is no limit on the price this supplier could bid and be called on to supply the needed 500 MWh. Even though there was a price cap in the California electricity market, something like this occurred because FERC required the California ISO to pay all units without Reliability Must Run (RMR) contracts that possessed local market power their bid price instead of providing the ISO with effective local market power mitigation, as it did for the eastern US ISOs.[2] Apparently, FERC did not recognize that local market power problems are endemic to all electricity markets with insufficient transmission capacity to support a competitive wholesale market at all locations in the network, regardless of the mechanism used to manage transmission congestion. As discussed earlier, a strong case can be made that all transmission networks in the US are insufficient to support a competitive wholesale market without effective local market power mitigation.

Although the California market is not unique in having a transmission network that is poorly suited to a wholesale electricity market, it is unique in being required to pay generation units (without RMR contracts) that possess substantial local market power their bid price to supply energy. In the PJM, New York and New England electricity markets, FERC allows the ISO to implement mitigation mechanisms that limit the potential harm to consumers from local market power. Under the scheme in PJM, the ISO implements a two-step process for local market power mitigation. It starts with the ISO operator's best estimate of system conditions for the following day, including the level of demand at each location in the network and the mix of available generation units in the control area, and implements a well-defined process for determining whether a unit possesses significant local market power. If this unit-level determination is affirmative, then the bids of this unit are mitigated to their variable cost plus a

[2] Reliability Must Run (RMR) contracts were contracts between the certain suppliers and the California ISO that allowed the ISO operators to call upon the unit to provide local reliability energy at a contractually determined price. Bushnell and Wolak (1999) discuss these contracts and a flaw in their initial design.

10 percent adder. Then the ISOs dispatch and price-setting algorithm is implemented with these cost-based bids in place of the unit's actual bids. All units are then paid the resulting market-clearing price. The other eastern US ISOs have different local market power mitigation measures, but none of them involve paying generation unit owners their bid price when they have local market power. Wolak (2002) discusses the PJM local market power mitigation mechanism in detail.

As an additional safeguard against local market power problems and system-wide market power problems during the first year of operation of the PJM market, all suppliers were required to submit regulated start-up, no load, and variable costs into the ISO's dispatch and price-setting algorithm. Generators were then paid the market-clearing price based on these cost-based bids. The ability of the ISO to dispatch and set prices using cost-based bids or mitigate to variable cost the bids of all units that it deems possess substantial local market power, prevents the wholesale market from causing significant harm to consumers.

It is important to emphasize that this safeguard is not without short-term inefficiencies and long-term costs. Because of the individual rationality constraint on firm behavior, one response to this mechanism by profit-maximizing firms owning a portfolio of generation units is to withhold the less expensive units in their portfolio from the market, either through planned or forced outages, so that higher-cost units will be called on to run more frequently and therefore set higher market-clearing prices earned by all of the units supplying energy in the firm's portfolio. A longer-term response from firms would be to reconfigure their generation units and how they operate them to shift costs from fixed costs to start-up, no load or variable costs in order to justify a higher cost-based bid, and set higher market prices. The problem of variable cost creep could be quite severe if the California crisis is at all representative of firm behavior.

Kolstad and Wolak (2003) find evidence that, during the summer of 2000, suppliers with some plants located inside the South Coast Air Quality Management District (SCAQMD) and others outside of the area used NOx emissions permit prices from the SCAQMD emission permit market to justify higher bids into the California ISO's real-time energy market from generation units with high NOx emissions rates. This strategy was very successful at raising prices in the CAISO's real-time energy market during the summer of 2000. We discuss specific problems with cost-based dispatch real-time markets that are specific to LACs later in the paper.

3.2.3. The ISO as a Market Operator versus a Market Participant

The ownership and mode of operation of the ISO provides another method for the market designer to implement regulatory safeguards. The most common model for an ISO in the US is as a non-profit public-benefit corporation. In most other countries, it is a government-owned corporate entity. Finally, in other markets it is a privately-owned regulated monopoly. ISOs also differ in terms of whether the regulator allows them to take a financial position in the market. The US takes the extreme view that the ISO is simply a neutral market and system operator. In England and Wales, National Grid operates the system in real-time and the regulator allows it to take a financial position in the market. In fact, the regulatory mechanism that sets National Grid's allowed revenues gives it strong incentives to reduce the real-time costs of operating the system.

The goal of forming the ISO as a non-profit or government-owned entity is to ensure that it will operate the grid in a non-discriminatory manner. However, if as is the case in all US markets and a number of countries around the world, retailers are subject to restrictions on their strategies for procuring energy and ancillary services to meet their load obligations by state public utilities commissions (PUCs) and/or national regulators, and suppliers do not face these same constraints, then retailers may not be able to take all of the actions necessary to protect themselves from high wholesale prices. In the US, many retailers complain that they are prohibited or limited by PUCs in the sorts of activities they can engage in relative to wholesale electricity suppliers. This is an artifact of suppliers being primarily regulated by FERC and retailers being primarily regulated by state PUCs.

This asymmetry between suppliers and retailers can create circumstances when an ISO that does not take a position in the market is forced to accept extremely high bids for small amounts of additional energy in the real-time market on a consistent basis. Even if the ISO believes that it can meet predictable incremental energy needs in real-time at a lower cost by purchasing more energy in the forward market than in the spot market, it is prohibited from doing so. Under the US approach, ISOs are only supposed to execute feasible trades among market participants.

National Grid Company (NGC) provides an alternative solution to this problem. It is a market participant charged with ensuring reliable transmission network operation in real-time.

Consequently, if NGC is able to find a cheaper way to procure the necessary real-time energy and ancillary services in the forward market, then the England and Wales regulator allows it to do so. The amount of revenues NGC is allowed to earn is based on how well it manages these real-time system operation costs. For this reason, NGC rarely accepts bids with very high energy prices, because it has the freedom to take a position in the forward market to protect against these high bid prices.

The upside of this model for the ISO is that it creates another sophisticated buyer of energy and ancillary services with the ability to take actions to reduce electricity prices. The downside is that over the long-term, retailers have less incentive to become sophisticated market participants, so that fewer load-serving entities and consumers make the investments required to become sophisticated players. The final downside is that unless the regulator is able to control what actions or financial positions in the market the ISO is able to take, adverse market outcomes beneficial to the ISO but harmful to producers and consumers could occur. The NGC ISO model requires a far more sophisticated and credible regulator than an independent market operator model, because the privately owned, profit-maximizing ISO is an essential feature of this scheme.

An interesting variation on this theme is a government-owned ISO is charged with operating the transmission network to facilitate a competitive wholesale market. This organizational form creates an ISO with more discretion to take positions in the wholesale market. However, given its ownership structure the ISO is less likely to engage in privately profitable actions that may cause consumer harm. This model for an ISO may result in a better balance between providing protection for consumers and reliable non-discriminatory access to the transmission network in LACs.

4. Challenges to Implementing a Competitive Wholesale Market Unique to LACs

This section describes seven challenges unique to LACs that are associated with implementing a competitive wholesale electricity market. The first is how to guarantee adequate energy to meet demand in a country where hydroelectric capacity is a significant portion of the country's generation mix. The second concerns the difficulty of establishing a forward market for electricity where privately owned suppliers can purchase long-term commitments to supply electricity that they can use to fund investment in new generating facilities. The third concerns the difficulty of establishing an independent regulator or regulatory body that has the legal authority and can accumulate the expertise necessary to be a credible and effective regulator. The fourth issue is the costs versus benefits of bid-based versus cost-based real-time dispatch of generation units and pricing of electricity and ancillary services. The fifth challenge is how to set the allowed revenues that regulated retailers can earn. The sixth challenge is appropriate mechanism to guarantee that adequate generation capacity is constructed to meet a rapidly growing demand for electricity. The final issue is the necessity of private ownership in the four segments of the industry—generation, transmission, distribution and retailing.

4.1. The Gambling with the Weather Problem and the Cost-of-Deficit Parameter

There are a number of potential pitfalls in attempting to establish a competitive wholesale market in a hydro-dominated system. The first is what I would like to refer to as the "gambling with the weather problem." The second is the problem of encouraging fuel diversity. The third relates to the problem of attracting new investment when much of the hydro capacity is owned by the government. A final issue—how to involve final demand in the wholesale market—is a source of potential benefits that should be easy for the market designer to exploit.

Because the only cost of producing electricity from a hydroelectric source is the opportunity cost of producing that energy during some other hour of the year, this creates opportunities for the market operator to take actions to impact this opportunity cost. Given the desire of politicians and regulators to keep the price of wholesale electricity as low as possible, the following strategy for dispatching generation in facilities is particularly tempting for the ISO to adopt in a hydro-dominated ESI.

A number of the hydro-based systems in Latin American use a stochastic dynamic programming approach to determine the opportunity cost of water each day. For example, in Chile the system operator uses a simplified stochastic dynamic programming model that uses as inputs forecasts of the water level of the largest hydro facility in the country, the regulated costs of fossil fuel facilities, and a number of other factors to arrive at an opportunity cost of water. Brazil uses a similar but more complex procedure because of its substantially larger control area and number of river basins to manage. An important input to all of these models is the cost-of-deficit parameter. This is the dollar per MWh cost of having insufficient water to meet the demand for electricity at the prevailing retail price. This cost-of-deficit parameter should be related to the price at which all consumers are willing to reduce their consumption. However, because the default retail price in most LACs is set independent of system conditions, the cost-of- deficit parameter is typically set by the regulator.

However, because of a desire to keep the price of wholesale power as low as possible, the regulator and government have difficulty in not setting this cost-of-deficit unrealistically low. For example, in Brazil the current value of the cost-of-deficit parameter is 350 Real/MWH, which is approximately \$100/MWh. Corresponding values of this parameter from developed countries are over \$1,000/MWh. For example, all of the eastern ISOs in the US have bid caps on their wholesale market of \$1,000/MWh, which can produce location prices above this level.

Figure 4 illustrates why LAC governments find it so tempting to set the cost-of-deficit parameter so low and why this leads to what appear to be energy shortages. For simplicity I consider only a single demand scenario and two possible supply scenarios—a low water year that occurs very infrequently and a high water year that occurs more frequently. Let $P_{Average}$ equal the wholesale price implicit in the fixed retail price of electricity. This price is set to recover the total costs of suppliers in both high and low water years. During high water years, fossil fuel facilities can be run less intensively, to the width of the bar labeled "Fossil Generation in High HydroYear," which is narrower than the box labeled "Fossil Generation in Low Hydro Year." The height of each of these boxes is the average total cost of supplying fossil fuel-based energy equal to the width of this box. The box labeled "Low Hydro Year" has a width equal to the production of hydro units in a low water year and height equal to the average total cost of producing this energy—the fixed costs of the hydro facilities divided by the output quantity. The box labeled "High Hydro Year" has width equal to the production of hydro units in a high water year and height equal to the average cost of supplying this energy. The reason the dashed line for the "Low Hydro Year" box is higher than the solid line for the "High Hydro Year" box is because the same fixed cost is relevant for both years, but more energy is produced from the hydro facilities in the high hydro year.

For simplicity we assume the demand for electricity is the same in high and low water years because the retail price is the same in high and low water years. This creates an apparent shortage condition in low hydro years that is driven purely by the fact that the price of electricity is not allowed to rise to the level necessary to equate available supply with demand during the very infrequently occurring low hydro years. Note that in both the high hydro years and low hydro years, the products $P_{Average}*Q_{Low}$ and $P_{Average}*Q_{High}$ both cover the total costs of producing hydro and fossil-fuel based electricity. Consequently, the government covers the total costs of supplying electricity under both hydro conditions at $P_{Average}$. However, because the cost-of-deficit is set too low, the average wholesale price is too low to prevent shortage periods during low hydro years. When low hydro years occur there is a shortage of electricity at $P_{Average}$ equal to Q_{High} - Q_{Low}.

This example shows that if the government or regulator is willing to gamble with the weather in the sense of setting a low cost-of-deficit parameter and betting on high hydro realizations, the government can achieve the very important political goal of low wholesale electricity prices and still maintain the financial health of the electricity supply industry. However, the downside of this unrealistically low cost-of-deficit parameter is that when low hydro conditions occur there will be a shortage of electricity at the prevailing fixed retail price.

The example also demonstrates an obvious but often ignored logical implication of using stochastic dynamic programming models to set the opportunity cost of water. If the cost-of-deficit parameter is set at a level that assumes shortages of water are not very costly, then periods of water shortages will occur. If the cost-of-deficit parameter is set at a level that assumes shortages of water are extremely costly, then periods of water shortages will rarely occur. However, an important corollary to these statements is that the price of electricity will, on average, be higher with a higher cost-of-deficit parameter.

Given the enormous economic and political costs associated with periods of shortages at the prevailing retail price, the obvious solution to this problem is to use a cost-of-deficit parameter in the stochastic dynamic programming model that reflects these enormous costs. This will ensure that wholesale prices rise during low hydro years and, if a significant number of consumers pay a retail price linked to the current wholesale price, these higher prices will allocate this lower supply of electricity to those who value it most and eliminate the apparent shortage of electricity.

Ideally, the cost-of-deficit parameter should be set equal to the highest willingness to pay for electricity among all consumers. Then, as the amount of available water declines and wholesale prices rise, consumers can choose not to purchase electricity. Under this scheme, any set of hydro conditions can be managed without having to resort to arbitrary curtailments of electricity supply to customers willing to pay for the scarcity value of this electricity.

Setting the cost-of-deficit parameter too low also has important implications for encouraging fuel diversity. Note that $P_{Average}$ is always below the average cost of producing fossil-fuel-based electricity. This means that unless the government pays a subsidy to the fossil-fuel-based entrants they cannot make enough money from selling in the wholesale market to justify their investment. Therefore, setting a realistic cost-of-deficit parameter is a necessary first step towards encouraging fuel diversity in hydroelectric-dominated LACs.

The fact that the hydroelectric capacity is government-owned exacerbates the problem of a new entrant earning sufficient revenues to justify his investment. Prospective new entrants are discouraged by the realization that the government does not want to raise the wholesale price during low hydro years, but is instead willing to declare a shortage period instead raising wholesale price if these hydro conditions arise. If the wholesale electricity price was allowed to allocate the available supply of electricity during low hydro years, the new entrant could recover his total costs plus an additional return to allow him to remain in the market during high hydro years when the average wholesale price is below the average total cost of a new fossil facility.

All of these problems with hydro-dominated systems underscore the importance of sending real-time price signals to final consumers. Different from the case of fossil fuel-based systems, in systems dominated by hydro production, demand response is often the only way to manage the system reliably. For fossil fuel-based systems, the regulator counts on higher prices to cause a supply response. Generation unit owners will run their machines more intensively or purchase more expensive input fuel on the spot market if electricity prices increase, either of which can eliminate a supply and demand imbalance. Unfortunately, as noted earlier, rainfall does not increase in response to higher prices in a hydro-dominated system. The only way to manage a supply and demand imbalance in this case is to send price signals to final consumers to reduce their consumption of energy.

Because price volatility in a hydro-dominated system tends to be seasonal, rather than hourly or weekly, there is little need for hourly meters in order to treat load and generation symmetrically in a hydro-dominated system. Retail prices could be adjusted prospectively on a monthly basis and existing metering technology could be used to implement retail prices that are responsive to real-time market conditions in the wholesale market—whether is it is a high or low hydro year.

The more an ESI is dominated by hydroelectric capacity, the more likely it is that system conditions will arise that cannot be managed without sending wholesale price signals to final consumers. The more an ESI is dominated by hydroelectric capacity, the more straightforward it is to send meaningful price signals to retail customers without additional retail infrastructure investments, because it is more likely that the spot price risk that must be managed is seasonal, so that existing monthly or even quarterly metering reading is sufficient to charge consumers prices that vary with wholesale market conditions. Each month or season, retailers can set an average wholesale rate implicit in the retail rate that reflects the best estimate of system conditions for the coming month or season. Setting retail prices in this manner will prevent eventual energy shortages because final customers will have a strong incentive to reduce their demand in response to current supply conditions.

4.2. Fostering a Forward Market for Electricity

As is emphasized in Section 3, spot electricity markets are extremely susceptible to the unilateral exercise of market power. The most effective way to limit the ability of suppliers to exercise market power in a spot market is to limit the amount of their production that is sold in this market. This is accomplished by providing strong incentives for suppliers to sign forward contracts with load-serving entities. There are two ways to cause existing suppliers to sign forward contracts: (1) make the forward market very attractive by setting high prices, or (2) drive down the expected spot market price because that is the relevant opportunity cost to suppliers that sign forward contracts. For new suppliers this second avenue is not available. If the spot market is unattractive they are less likely to build new capacity.

Consequently, the challenge to LACs is how to develop a forward market that will support the construction of new capacity. LACs also have the additional problem that negotiations between load-serving entities and electricity producers may result in these two parties coordinating to raise the purchase price of the contract at the expense of final consumers. For example, in Brazil many of the retailers have affiliates that also own generation units, so there is a strong incentive for the retailer to purchase energy from its generation affiliate at an inflated price that is passed on to the retailer's customers.

One way to deal with this problem is to set up periodic anonymous auctions for standardized load shapes for delivery at specific locations in the transmission network. For example, one auction could be for supplying 1 MWh, twenty-four hours a day for five years, at a pre-specified location in the network, starting next year. All suppliers would submit their bid price and quantity for supplying this standardized load shape. The regulator would then determine the market-clearing price and how much each supplier won. These standardized contracts could then be traded on the secondary market and used to finance new investments in generation capacity.

The advantage of these periodic formal auctions of standardized contracts is that they would produce prices that could be used to set reference levels for the wholesale price component of retail electricity prices. In addition, because most LACs are dominated by hydro capacity, the primary source of price variation is across seasons, so there may be less need for a large number of standardized contracts. The only important difference in electricity prices may be across locations as a result of congestion, but not over time at that location, because the opportunity cost of water implies equalizing electricity prices at the same location over time.

Although the requirement to sell only standardized contracts may raise the cost of supplying forward contracts for electricity, the increased competitive pressures firms face, because these contracts are sold in an anonymous auction according to well-defined market rules open to all suppliers, should make the auction sufficiently competitive that bidders have very little opportunity to exercise any significant market power.

4.3. Establishing an Independent Regulator

Establishing a credible and effective regulatory process is an ongoing process. The US has been involved in this process for almost 100 years at the state and federal level for electricity, telecommunications and natural gas, and significant mistakes are still being made. Regulation is necessarily imperfect, for the simple reason that the regulator can never know as much as the firm does about its production process, demand or service territory, and these factors are continually changing.

Regulation also requires turning vague statements such as “just and reasonable prices” into operational concepts. Over the past 70 years there have been a number of legal disputes in the United States that have further refined this concept, but the process is far from over. It is important to recognize the need for continuous change in any regulatory process.

An important lesson from the transition to a competitive wholesale market regime in the United States is that the regulatory mechanisms that worked in the former vertically integrated monopoly regime are inappropriate for the new competitive wholesale market regime. In the former regime, the process of setting just and reasonable prices focused on collecting and analyzing balance sheet information from market participants according to well-defined accounting techniques. Consequently, the rate-setting process involved primarily lawyers managing the process and accountants implementing procedures to compute prices. In the competitive market regime, the US process is concerned with setting just and reasonable market rules, mechanisms for compensating market participants that will result in just and reasonable market outcomes. This means that if firms respect their individual rationality constraints, market outcomes will yield prices that do not harm consumers yet still recover the production costs of all suppliers. This is a process that requires a substantially more sophisticated regulator, because it must be able to anticipate the likely response of market participants to any market rule changes or any regulatory intervention it might undertake. This is an extremely difficult task, but one way to make it easier is to enlist the help of the public.

The regulator can best enlist the help of the public in obtaining just and reasonable market outcomes by recognizing that the primary role of regulation is information provision. Perhaps the most cost-effective form of regulation in a market environment is sunshine regulation—shining the light of public scrutiny on the behavior of certain market participants or the individual profitability or cost to system reliability and market efficiency of certain market participant actions. This regulatory strategy underscores the importance of disseminating the raw data on market outcomes and market participant behavior to the public as well as the need to establish a set of standardized measures of market performance that are disseminated on a regular basis.

There is a natural dividing point between data that should be made available to the public and data that should remain confidential. Specifically, any information submitted to the spot market and system operator or produced by these entities should be made available to the public. Information on supplier and generation unit-specific bids to the day-ahead and real-time energy markets should be made available, as should data on the actual output of each generation units by owner and withdrawals of electricity at each location in the transmission network. Information on the amount of transmission capacity available on day-ahead scheduled and real-time basis and the scheduled and real-time energy flows along the various transmission paths should also be made available. Finally, if the market is dispatched on the basis of cost-based bids, these bids should be disclosed to the public. The logic underlying this information disclosure requirement is that day-ahead and real-time market operation should be transparent as possible to all market participants. Any market participant should be able to replicate the actual day-ahead and spot market outcomes down to the specific generation unit and take-out point for load-serving entities.

This information should be released to the public with the minimum possible time lag with supplier-specific identifiers for each market participant. There are a number of reasons why data release should be immediate. The first is that it will be very difficult for market participants to make the best use of this information if it not immediately available to them. Although there may be concerns that releasing this information with a short time lag may help suppliers to coordinate their actions to raise market prices, this concern is much less relevant if the day-ahead and real-time markets are operated using cost-based bids. If the market is bid-based, this may be a reason to delay data release up to one month from the actual date of the market outcomes. Immediate data release helps solve another problem in developing countries—establishing credibility of the regulatory process. If the regulator is able to make its case in the court of public opinion using data that is available to all interested parties, it is less likely to be influenced by the political process, because any claim it might make can be verified by interested parties. A pre-commitment to immediate data release also provides strong incentives for market participants to treat the day-ahead and real-time energy markets as balancing markets, rather than markets where any sizeable amount of net energy is bought or sold. In this way it provides an additional incentive for the establishment of an active forward market for energy. Moreover, to the extent that the operation of short-term energy markets that forward contracts clear against are better understood by all market participants, the smaller will be the risk premium that each side of demands to sign a financial arrangement for delivery of electricity at some date in the future. If a supplier or load-serving entity is unable to forecast spot market outcomes accurately, they will demand a risk premium to engage in forward contracts to compensate them for this risk. Consequently, it is in the interest of all parties involved for these short-term markets to be as transparent as possible. Another important reason for immediate data public data release is that it reduces the barriers to new entry. If the short-term energy markets are well-understood and transparent to a large number of potential and actual market participants, this will reduce the cost to a prospective new entrant to gathering the necessary data to perform a feasibility study for determine whether and where to enter the industry. In addition, if there are standardized measures of market performance computed on a regular basis available to the public, this should further reduce the barriers to new entry, which will further increase the competitiveness of the energy and capacity markets. A final reason for immediate public data release is that it should improve the process of determining the feasibility of transmission upgrades. One could imagine the ISO computing and dissemination on the causes and costs of congestion on each transmission path as a mechanism for encouraging market participants and regulators to determine whether a local demand needs is served at least cost by a transmission upgrade or new generation entry. This data release policy for the regulatory is consistent with the view that the day-ahead and real-time markets are primarily mechanisms for ensuring system reliability and market efficiency, something all market participants should have a common interest in.

The forward energy market, defined as all markets that clear more than one day in advance of delivery, should be a primary market where energy is purchased and sold. Information on these transactions should remain confidential. There is no system reliability or spot market efficiency rationale for requiring this information to be publicly disclosed.

Specifically, the system or spot market operator does not need to know these forward market positions in order to reliably and efficiently operate the system on a day-ahead and real-time basis. Therefore, market participants should be free to take financial positions in these markets that they are able to keep confidential. This is not to say that under certain circumstances the regulator should not be able to obtain this information from market participants, only that it should not be required to be released to the public on a regular basis.

There are two other reasons for this dichotomy in the regulator's data release policy. The first relates to point made in Wolak (2003c) that the bid curve of a given supplier conveys little information about that supplier's variable cost of producing electricity unless the researcher knows the supplier's forward market position. Consequently, requiring a supplier to disclose bids or production levels should not allow current or future market participants to infer the nature of a supplier's production costs or the prices it paid for input fuels. In addition, because forward market positions are purely financial, it is difficult determine the true forward financial position of a specific market participant because it may have signed another financial contract with an affiliate company that completely unwinds any forward financial position it might report to its regulator. For example, even if a given company is required to report all forward contract positions it has signed to the regulator, this company may have signed forward contracts with a parent or affiliate company that the regulator has no jurisdiction over. This parent or affiliate company could then split off the components of this forward contract position to other entities to completely unwind the forward market position that was reported to the regulator by the market participant. This possibility limits the usefulness of forward market information to the regulator in a market where many of the suppliers are a part of large, diversified companies, as is the case in many companies. This makes the case for the regulator collecting forward market information much less compelling. In contrast, because day-ahead and real-time market information determines how specific generation units will operate and how much energy load-serving entities will ultimately deliver to final consumers, this issue is not relevant for data from the short-term energy markets.

The final issue on the design of a regulatory oversight process is how to establish credibility and expertise. Credibility is similar to reputation, and the only way to establish a reputation for sound decisions is to always make sound decisions. This implies that the regulator must slowly build up its reputation by making decisions that it believes can withstand judicial review. If the regulator is always upheld on judicial review, this establishes a reputation that the regulator can exploit over time. By starting with smaller decisions that can withstand judicial review, the regulator builds a reputation for substantial expertise that can be exploited to support some decisions in the future requiring more discretion by the regulator. As noted above, making these decisions based on publicly available data is another way to establish credibility.

A second way to establish regulatory credibility is to perform analyses in accordance with international standards for market monitoring and oversight. There is a generally acknowledged set of market performance indices used in most developed countries. Competitive benchmark pricing analyses, as discussed in Borenstein, Bushnell and Wolak (2002), are now performed in all US markets and a number of international markets. Computing the frequency that large suppliers are pivotal in an energy or ancillary services market is also a part of the market monitoring process in a number of US and international markets. All of the market monitors in US produce annual summaries of market performance that share a surprising amount of agreement on the market performance measures reported. LAC regulators should adopt these similar reporting standards as a way to increase their expertise and credibility.

Another way to use international market monitoring standards to improve the credibility and expertise of the domestic regulatory body is to establish an independent market advisory committee. This committee would be composed of three to four international experts on electricity market design, monitoring and regulation. Consistent with the goal of sunshine regulation described above, the primary role of this committee would be information provision. It would be charged with monitoring the performance of the market and the market and system operator. The committee would have no formal decision-making power or ability to impose penalties or sanctions on specific market participants. However, it would have the ability to issue opinions to the public on specific issues the regulator or system operator would like it to address, or that it decides to address. With the assistance of the system and market operator, it would also prepare periodic reports on the performance of the ESI that would be publicly available. For this reason it should have access to the all market data and any confidential data the regulator and market operators are able to obtain.

There are a variety of ways this independent committee can improve the credibility and effectiveness of the regulatory process. By issuing an opinion on a contentious issue that provides perspective on how this issue has been resolved internationally, the independent committee can make it more difficult for the government and judiciary to go against a decision by the regulator that is consistent with precedents from other countries around the world. Because the committee would have the ability to issue opinions on any issue it deems worthy of further study, it would have the ability to point out market design flaws the regulator may have missed either on purpose or because of political pressure to ignore them. This committee would be expected to meet approximately every other month but be available for telephone consultations with the regulator and its staff on a more regular basis. These interactions among participants will increase the transfer of international expertise in market monitoring and regulation to the domestic regulator. Because this committee has no legal power besides the ability to issue opinions and demand and receive information that it analyzes to formulate these opinions, it can be a very effective neutral forum for resolving many controversial issues between the stakeholders. Given the difficulty that virtually all LACs have had in establishing a credible regulatory mechanism, the potential benefits of forming such a committee, at least for the first five years following the reform, appear to outweigh the costs for most all LACs.

4.4. Bid-Based versus Cost-Based Dispatch and Pricing

One lesson from ESI reform in LACs that has not received much notice from developed countries is the use of cost-based dispatch and pricing in short-term energy and reserve markets. With the exception of the single transitional year in PJM in 1998 discussed earlier, no developed countries have implemented cost-based dispatch or pricing schemes. All of these markets allow generation unit owners to submit bids in the form of either simple supply curves that give each unit's willingness to supply energy as a function of the market price or multi-part bids, usually with a start-up bid, no-load bid and a willingness to supply curve bid as a function of the market price. As noted previously, bid-based dispatch and pricing has two obvious disadvantages, particularly in markets with insufficient transmission capacity to face all suppliers with enough competition at all locations in the network. The first is system-wide market power, which is the ability to raise the system-wide price by bidding a willingness to supply curve that exceeds a unit's marginal cost curve at the level of output from the unit that the firm expects to sell in the market. The second problem is local market power, which occurs when a unit or set of units faces an extremely inelastic residual demand curve (usually perfectly inelastic) because of transmission constraints into a geographic area. Under both of these circumstances, prices in a bid-based market can be expected to be substantially in excess of the marginal cost of the highest cost unit operating in the market for the case of system-wide market power, and in a smaller geographic area for the case of local market power.

Starting with Chile, virtually all LACs, with the exception of Colombia, dispatch and set prices using regulated unit-level costs as opposed to bids. This is an important safeguard that allows a country to avoid the enormous expense of setting up a bid-based dispatch and pricing process. The cost-based approach also allows the ISO to avoid the time and expense of formulating a local market power mitigation mechanism, which is essential in an ESI that does not have a transmission network that can support a competitive wholesale market, as is the case in all LACs. The quality of the transmission network in most LACs provides another argument in favor of a cost-based dispatch and pricing mechanism.

One argument in favor of the bid-based system is that it allows hydroelectric unit owners to manage their water more efficiently. They can bid their willingness to supply electricity, or equivalently water, and raise the price of electricity when they believe scarcity conditions are more likely. These higher prices will reduce the likelihood of electricity shortages because fossil fuel units will be run more intensively, much earlier in a low hydro year. However, unless final consumers actually pay these higher prices they will continue to consume at the levels they would in any other year with the same weather conditions and fixed retail price. This will cause a so-called "shortage period" that is not a shortage of energy to meet the demand at any finite price, but an unwillingness to let the retail price rise to the level necessary to allocate the available water.

Returning to our example in Figure 5, the most likely outcome in a bid-based market during a low water year is that the wholesale prices rise because hydroelectric capacity owners bid higher prices to conserve their water. However, if the retail price is maintained at $P_{Average}$, there is still a shortage of energy, because retail demand stays at Q_{High} and there is not enough water to meet this demand. Because hydroelectric suppliers bidding higher prices will cause fossil fuel generation units to operate more frequently than they would with lower wholesale prices, the amount of energy available in a low hydro year should be greater than it would be if the wholesale price were not allowed to rise. However, if there is still a shortfall of energy relative to the demand at $P_{Average}$ with these fossil fuel units operating more intensively, the retail price must be increased above this level or rationing will still be necessary. There is no way to avoid the basic reality that demand must be the marginal supplier of additional negawatts of electricity during extremely low water conditions, regardless of the decision to use a bid-based or cost-based spot market.

There is also a more straightforward way to ensure that the fossil fuel units are used as efficiently as possible during low water conditions. Raise the cost-of-deficit parameter to a level that is at least equal to the highest individual willingness to pay for electricity of all consumers in the system. This cost-of-deficit parameter will cause fossil fuel facilities to be operated more intensively sufficiently far in advance to provide at least as good of a hedge against water shortages as allowing hydroelectric suppliers to bid their willingness to supply energy. However, this scheme does not have either the system-wide or local market power risk of suppliers creating an artificial scarcity to raise the price of electricity.

Consequently, particularly for hydro-dependent systems, shortages must ultimately be managed by allowing loads to be the marginal supplier of negawatts, meaning that the willingness of consumers facing retail prices that vary with real-time system conditions to reduce their purchases is the only way to solve the demand and supply imbalance during certain system conditions. Because higher prices do not cause more rainfall, there is a sufficiently low water level where demand reduction is the only way to maintain system balance. Increasing the retail prices consumers pay during these circumstances is the only way to make consumers more willing to supply the necessary negawatts.

Cost-based markets have an additional advantage that reinforces the goal of fostering active forward markets for energy. In particular, assuming that the regulator follows the data release policies given above, suppliers will have a much easier time forecasting spot electricity prices in a cost-based market versus a bid-based market, because they will not have forecast the ability of suppliers to exercise market power through their bids into the spot market. Although it is extremely difficult to forecast the ability of suppliers to exercise market power by bidding into the spot market, as the California and New Zealand markets vividly demonstrate, the ability of suppliers to exercise unilateral market power is a major determinant of the mean and variance of prices in short-term electricity markets. A cost-based market eliminates the variation in spot electricity prices that occurs from suppliers bidding to raise or lower these prices. This should reduce the cost of suppliers and load-serving entities signing forward contracts, because both sides will have less uncertainty about the time path of spot prices over the duration of the contract. Both parties can compute forecasts of future spot prices using the publicly available market outcomes data along with the cost-based dispatch algorithm.

Although suppliers are constrained to bid their regulated costs in a cost-based market, this does not eliminate the incentive or ability of privately owned suppliers to exercise market power. Consistent with the discussion of Section 2, these attempts to exercise market power simply take a different form. Specifically, suppliers can now be expected to attempt to raise their regulated costs of production that enter the dispatch process. Consequently, a necessary pre-condition for a cost-based market is that the regulator must have in place a mechanism for determining whether the production costs a supplier reports have been prudently incurred. As noted in Wolak (2003b) with respect to the soft price cap implemented by the FERC in the California market in December 2000, a cost-justification process that does not impose a prudency review on how costs are incurred effectively provides no protection against the exercise of unilateral market power, because suppliers can make fuel costs and other input costs equal whatever level they would like bid through transactions with affiliate companies, so that a cost-based market becomes equivalent to a bid-based market.

The experience of California with the soft price is instructive in this regard. Many of the electricity suppliers in the California market had natural gas supply affiliates. The FERC soft cap policy allowed a supplier that could cost-justify its bid above $150/MWh to be paid as bid for their energy. The March 2003 FERC staff report on western US energy markets found substantial evidence that suppliers inflated the natural gas prices reported to FERC as justification for bids above the $150/MWh soft price cap.[3] The solution to the problem of

[3] "Final Report on Price Manipulation in Western Markets: Fact Finding Investigation of Potential Manipulation of Electric and Natural Gas Prices," Docket No. PA02-2-000, March 2003, Prepared by Staff of Federal Energy Regulatory Commission.

inflated input fuel prices recommended by the FERC staff is to use the spot natural gas price at Henry Hub, a very liquid spot market for natural gas in Louisiana, plus the FERC-regulated cost of transporting gas from Henry Hub to California as the relevant input fuel price for natural gas-fired facilities.

Similar problems arise with respect to cost-based dispatch electricity markets. The regulator cannot simply accept without review a supplier's claimed input costs. Instead, the regulator must establish an administrative procedure for determining the input costs that are consistent with prudent purchasing behavior. These prudently incurred input costs should then be used to set each supplier's variable cost for the cost-based electricity spot market. This requirement for the regulator to set standards for prudently incurred input costs requires the regulator to monitor the performance of input fuel markets and other input markets in order to make a determination of whether the input costs that a supplier claims are in fact appropriate to include in its regulated costs.

Another approach builds on the mechanism recommended by the FERC staff for the western US electricity market. The regulator would construct indices of input fuel prices based on forward contracts and spot purchases that it was able to verify were reflective of actual competitive conditions in these markets. These indices would then be used to set the variable costs of each generation unit that subsequently entered the cost-based dispatch process. There are a number of ways to structure the administrative process for validating a supplier's claimed variable costs. The major point that I would like to emphasize is that such a process is an essential feature of a cost-based electricity market.

This discussion underscores the major lesson from Section 2, that regulatory mechanisms and market mechanisms both have their strengths and weaknesses. Cost-based electricity markets have the increased regulatory burden associated with setting a standard for prudently incurred input costs for each generation unit. To set these regulated costs of production, the regulator must also engage in more extensive data collection and analysis of market performance for input fuel and other inputs to the electricity production process. Bid-based dispatch markets have the problems with system-wide and local market power discussed above. Although cost-based dispatch and pricing also has the long-term efficiency costs discussed in Section 3.2, given the enormous expense of setting up a bid-based spot market, the enormous potential downside of such a market in terms of system-wide and local market power, and the relatively small potential gain from a bid-based dispatch and pricing in an ESI without active demand-side participation in the wholesale market, a superior strategy for LACs is cost-based dispatch and pricing for the foreseeable future. The potential benefits of bid-based dispatch and pricing and the huge uncertainty associated with these benefits in an ESI with a less-than-ideal transmission network and nascent regulatory framework makes cost-based dispatch even more attractive in spite of the regulatory challenges associated with determining each unit's cost of production.

4.5. Regulating Default Retail Rates

Determining the wholesale power cost component of the regulated retail electricity price is a challenging task even for developed countries. The regulator must effectively run an energy trading firm to compute the best possible estimate of the least cost combinations of long-term contracts, medium-term contracts, spot purchases and other hedging instrument purchases. Because all transactions but spot purchases are made through bilateral negotiations, it is difficult for the regulator to know which transaction prices are legitimate and which are not. For example, if the buyer and seller of electricity have some sort of financial relationship, the buyer may be willing to buy at an inflated price if these prices can be passed on to final consumers.

Consequently, the challenge for the regulator is how to set up a mechanism that yields the most useful information about the opportunity cost of power delivered on a given date and at a given location that can used to set the default rate for an electricity retailer. This mechanism must guard against affiliate-dealing to raise the prices that consumers must pay. Kolstad and Wolak (2003) present an example of this sort of problem in the California market with respect to NOx permit prices. As discussed earlier, suppliers with units located in and outside of SCAQMD, the geographic area covered by the emissions permit market, were willing to pay inflated prices for these permits because the permits enhanced the ability of firms to raise wholesale electricity prices. Kolstad and Wolak (2003) suggest a mechanism that can be used to limit the ability of suppliers to use NOx permits to raise electricity prices. They argue that rather than allow suppliers to engage in bilateral NOx permit trades at any mutually agreed upon time, SCAQMD should run periodic anonymous auctions in which suppliers bid for the right to buy or sell these permits. A single market-clearing price would then be set for all of the permits sold.

This price is more difficult to move because it is the result of the intersection of the aggregate demand bid curve for these permits with the aggregate supply curve.

This solution could be used by an LAC regulator to set the wholesale energy price implicit in the default retail price. The regulator would define standardized forward contract products for specific locations and durations and then require retailers to use the anonymous auction market operated by the regulator to purchase the forward energy requirements (as determined by the regulator) necessary to meet its default provider obligation. In a steady state, this process would produce a portfolio of forward prices for delivery at a given location and date in the future. The prices for, say, May 2005 delivery implicit in contracts starting in each month of 2004 could be used to determine the wholesale price component of the default provider's retail rate. For example, one scheme could use a pre-specified weighted average of the May 2005 forward prices that result from these auctions as the wholesale cost component implicit in all regulated retail sales for May 2005.

This would accomplish two goals for the LAC regulator. First, it would limit the ability of retailers to self-deal with their generation affiliates. Second, it would allow the regulator to avoid the extremely difficult task of setting the level of the wholesale price component of the default provider retail price. Variations on this theme could also be implemented, but the basic idea is to replace negotiated bilateral purchases with periodic formal forward markets for electricity as a mechanism for retailers to purchase the forward market obligations necessary to meet their default provider obligations.

Another mechanism that has been suggested to solve this problem is to establish a government-owned single buyer of wholesale electricity that subsequently sells this wholesale energy to electricity retailers. Although at first glance this solution seems to have much to recommend it, because it pits a single state-owned buyer against many sellers, this approach is nothing more than replacing one state-owned monopolist with another state-owned monopolist. The option of the former state-owned monopoly electricity supply company to purchase all its energy from electricity producers as a single buyer existed under the former vertically integrated monopoly regime. The state-owned monopolist could have held periodic negotiations with suppliers to purchase its electricity needs for a certain number of years into the future, as a single buyer would do. However, the former state-owned electricity supplier found it optimal to instead construct and operate power plants. Consequently, the single buyer alternative can be thought of as implementing a solution that was available in the former monopoly regime, but not found to be optimal. For this reason, a single buyer solution is likely to yield market outcomes that are more harmful to electricity consumers than the former vertically integrated monopoly regime.

It is important to stress that the single buyer model has limited incentives to minimize wholesale energy procurement costs for the same reason that the former state-owned ESI had limited incentives to minimize the cost of supplying wholesale electricity. Both of these entities are state-owned, so we do not expect either to have strong incentives to minimize wholesale energy costs. Moreover, building on the discussion in Section 2, having a large number of buyers of wholesale electricity, each competing to supply final consumers with electricity, should lead to lower wholesale prices than a state-owned single buyer. Rewarding the load-serving entities that procure wholesale power at least cost with the most retail load, provides very strong incentives for least-cost procurement of wholesale power. In contrast, the single buyer does not lose any customers if it does not obtain the lowest possible price for wholesale electricity, because it is the monopoly buyer of wholesale power. Moreover, it is also very likely that the state-owned single buyer will pursue other goals besides least-cost procurement of wholesale electricity because of the political pressures its management faces as a result of this entity's government ownership.

Consequently, despite apparent attractiveness, by applying the logic presented in Section 2, the single buyer model is seen to be an available option not chosen by the former vertically integrated monopoly regime. In this regard, it is important to emphasize another major lesson from Section 2: unless the new market design changes the incentives faced by market participants, their behavior will not change and market outcomes will not change. In the present case, the single buyer still faces limited incentives to make the incurred cost of purchasing wholesale power equal the minimum cost of procuring wholesale power for many of the same reasons that the vertically integrated state-owned monopolist had limited incentives to make the incurred cost of producing wholesale energy equal the minimum cost of producing wholesale energy.

4.6. Capacity Adequacy in Rapidly Growing ESIs

A major concern in LACs is whether there will be sufficient new generation capacity built in a timely manner to meet the rapidly growing demand for electricity in these countries. Capacity payments have been proposed to encourage the necessary new investment. Although I have no quarrel with the argument that paying suppliers a $/MW payment per day or per month for installed generation capacity provides incentives for the construction of new generation capacity, the more important question is whether these capacity payment schemes are the least-cost way to ensure future capacity adequacy. The following logic demonstrates there are substantially lower-cost ways to achieve capacity adequacy.

The first shortcoming of capacity payment mechanisms is that it is unclear what product is actually being purchased. The analogy to a capacity payment in other industries would be monthly payments to factories or merchants for simply being in existence. It is hard to find other market environments where this kind of payment is made. Instead, the supplier is only paid for the products it sells, not merely for being in existence. Although a number of advocates of capacity payments attempt to call them capacity markets, this is really a misnomer. As all Eastern United States ISOs with capacity payment schemes have discovered, it is virtually impossible to operate a market for installed capacity. The marginal cost of supplying additional capacity is zero if there is adequate capacity in the control area and effectively infinite (in the short-run) if there is insufficient capacity. Moreover, because the demand for capacity is perfectly inelastic—load-serving entities are typically required to buy some multiple of their peak demand in installed capacity—this can present severe market power problems.[4] Specifically, if a supplier knows that its generation capacity is needed to meet the aggregate installed capacity demand (it is pivotal), this supplier can set the price of installed capacity at whatever level it would like. For this reason, all of the capacity "markets" in the eastern US have price caps or maximum payments load-serving entities must make if they do not satisfy their installed capacity requirements. Because what are often called capacity markets are in reality capacity payment schemes, this raises the important problem of how to set the appropriate level for the capacity payment. The regulator is virtually certain to get this $/MW per month

[4] The PJM and ISO-New England ISOs have had substantial market power problems in their capacity "markets."

payment wrong, in the sense of providing incentives for too much or too little generation capacity.

To understand the problems associated with setting the level of the capacity payment it is helpful to delve deeper into the rationale for capacity payments. Unloaded generation capacity is needed in real-time to meet unexpected surges in demand and to ramp up quickly if a power plant is forced out or a transmission line is unexpectedly out. Consequently, in order to meet a peak demand of 10,000 MW with some pre-specified level of reliability, more than 10,000 MW of generation capacity is needed. There is a more efficient way to pay for the necessary operating reserve capacity than paying all generators a fixed $/MW per month. All wholesale electricity markets in the US operate ancillary services markets to procure this standby generation capacity. Suppliers bid their willingness to supply energy and ancillary services and the market operator determines the market-clearing price for each of these services, along with the price of energy. Consequently, a supplier that is providing standby capacity is paid a $/MW charge for each hour it holds 1 MW of unloaded capacity providing operating reserve. The ancillary services market only pays those generators providing standby capacity for the service they provide. Units in the control area not providing energy or standby capacity service during an hour do not receive ancillary services payments.

This ancillary services market mechanism directs capacity payments to those units that provide real-time reserve capacity. These tend to be the high-cost, fast-start units. Although these units only recover their variable costs or slightly more by providing energy, they are able to obtain full cost recovery over the year because they provide standby capacity for virtually all of the remaining hours of the year when they do not operate. Low variable cost units, which tend to run the majority of hours, can obtain full cost recovery from selling energy because their variable cost is below the market price the vast majority of hours of the year. There is no need to pay these units a capacity payment to keep them in the market. Consequently, there is no need to make capacity payments in a market that pays for all of the real-time operating reserve that the system operator uses.

A final problem with capacity payments is that they do not, in general, provide an incentive for units to be available when the system operator needs them. In other words, capacity payments do very little to eliminate the incentive that suppliers have to withhold capacity from the spot market. As discussed in Wolak (2003b), the cause of the California electricity crisis was not insufficient generation capacity relative to the level of demand, but insufficient incentives for suppliers to make all of the capacity they owned available to the spot market. Therefore, the problem of capacity adequacy should be reformulated as ensuring that load-serving entities have purchased sufficient quantities of energy in the forward market to ensure enough energy is made available to the short-term market for the operators to be able to reliability operate the system in real-time. For example, suppose a load-serving entity buys 95 percent of its expected demand for the coming year in the forward market. For the sellers of these forward commitments to be confident they can supply the required MWh, more generation capacity must exist than the maximum hourly demand in the contract, because of the generation and transmission availability uncertainties described above.

Rather than thinking about the capacity adequacy in terms of having sufficient generation capacity constructed inside the control area of the ISO, the question should be recast as that of insuring energy adequacy. Specifically, do all of the load-serving entities have sufficient quantities of forward energy commitments of a long enough duration into the future to be confident they can meet their future energy needs with these contracts and spot market purchases? Therefore, load-serving entities should not be concerned with the question of the existence of sufficient generation capacity constructed in the ISO's control area to meet future demand. Instead, they should focus on the question of purchasing enough energy far enough in advance of delivery to be able to meet their real-time load obligations with a desired level of reliability, and leave the decision of how much capacity to build to achieve this level of reliability to the sellers of these forward contracts. If the sellers believe additional generation capacity is needed to meet the forward energy obligations they have sold, they have a strong incentive to construct these facilities or bear the large financial risk of being short relative to their forward energy obligations in the short-term energy markets.

This discussion of capacity markets can be summarized as follows. Having adequate capacity to meet demand does not guarantee that this demand will ultimately be served. Having purchased sufficient energy in the forward market far enough in advance of delivery guarantees the load-serving entity is hedged against spot price risk. Consequently, the major challenge associated with capacity adequacy is providing load-serving entities with strong incentives to accurately forecast their future energy needs and buy them far enough in advance real-time delivery so that suppliers can construct the necessary generation facilities to meet these future energy obligations. Finally, as noted above, capacity payments do little to solve this future energy adequacy problem.

4.7. Government versus Private Participation in the ESI

Section 2 compared the incentives for firm operation provided by government versus private ownership. I now use that analysis to provide recommendations for what portions of the ESI in an LAC should be privately owned and which can remain in government hands initially and even indefinitely. For this discussion it is useful to divide the electricity supply industry into four segments: (1) generation, (2) transmission, (3) distribution and (4) retailing.

Because the technology of producing transmission services for a given geographic area dramatically favors supply by a single firm, the price generators and retailers pay for accessing the transmission network must be set by an administrative process, whether or not the network is government-owned or privately owned. In addition, the transmission network access charge usually makes up between 10 percent and 15 percent of the retail price of electricity. The largest cost associated with providing transmission services after the network has been built is for maintaining the network. These last two points argue in favor of the view that cost differences due to productive inefficiency differences caused by government versus private ownership will not be a very large fraction of the retail price of electricity.

There are also potential benefits, particularly in developing countries, associated with keeping the transmission network in government hands. Obtaining rights-of-way and environmental approval for transmission expansion may be less costly if the government rather than a private investor owns the transmission network. Consequently, even though the incurred cost of operating the network may be greater for a government-owned firm (because of the associated productive inefficiencies resulting from government ownership), this may be offset by the lower transaction costs associated with expanding the transmission network. This logic suggests that keeping the transmission network in government hands would not be harmful to the formation of a competitive market and may even result in lower delivered prices to consumers.

The argument for government versus private-ownership of the distribution network is similar, but slightly stronger in favor of private ownership. The cost of the distribution network is a larger fraction of the price of electricity than is transmission. This means that increased inefficiencies due to government ownership could mean higher prices under government versus private ownership. The siting advantage for government ownership is also less relevant for distribution because there is rarely an alternative path for the distribution wire to a customer's house, but there are often many possible routes for bulk transmission lines. Also, if consumers want electricity they must be willing to tolerate the construction of a local distribution network.

Particularly, for the case of LACs, where the percentage of the population with access to electricity is significantly less than 100 percent, privately-owned profit-maximizing firms should have a stronger incentive to expand their distribution networks to serve these areas than government-owned firms. Consequently, assuming a properly designed mechanism for setting the regulated price for accessing the local distribution network, private ownership should be favored.

The case for privately owned retailers is very strong. In many developing countries the government-owned company often finds it impossible to get a substantial fraction of their customers to pay for electricity. Moreover, the company often finds it difficult to disconnect customers that do not pay because they are politically powerful. Turning ownership over to private investors provides much stronger incentives for retailers to get consumers to pay their bills. Any money the retailer does not collect from consumers must come out of the pockets of the firm's owners. In contrast, government-owned firms can fund this shortfall from tax revenue. Private ownership also makes it easier for the retailer to disconnect from the network customers who do not pay their bills. For all of these reasons, retailers should be privately owned as soon as possible.

In most LACs, a large fraction of the generation capacity is initially owned by the government. Clearly, reducing the government's ownership share will give private investors greater confidence in their ability to invest in new capacity and earn sufficient revenues from electricity sales to earn a reasonable return on the investment, because they know that more generation capacity is owned by entities that face the same budget constraints as they do. They are likely to be less fearful that their investments might be expropriated by a future government that releases water from its hydroelectric facilities too quickly in order to reduce wholesale electricity prices for political reasons.

On the other hand, private ownership puts generation capacity in the hands of entities with strong incentives to use it to raise wholesale electricity prices. However, so long as system dispatch and pricing are based on regulated costs, this risk should be less of a cause for concern.

A strong point against immediate plant divestiture to private owners is that the capacity may be sold at a discount because the buyers are not confident that the government will support the new competitive wholesale market. By keeping the capacity in government hands and operating under the new regime for a couple of years, investors should gain greater confidence in the government's support of the new wholesale market regime. The conclusion for the generation sector is private ownership as rapidly possible, consistent with obtaining a reasonable price for the government's generation assets.

There is an additional argument in favor of private ownership in both electricity production and retailing. This has to do with the liquidation of insolvent firms. One source of potential benefits from a market mechanism is the fact that inefficient suppliers are forced to exit the industry. One measure of the success of a re-structuring process is the extent to which inefficient suppliers exit and efficient suppliers grow. In a market with privately-owned companies, this process is relatively straightforward as long as the country has well-defined bankruptcy laws that are enforced. If a firm reaches the point where it is unable to pay its creditors, they can demand that the company liquidate its assets. Usually, companies in poor financial condition forecast this endpoint and attempt to exit the industry in a manner that maximizes the value of the firm's assets. Government-owned firms complicate this process because governments are extremely reluctant to allow the companies they own to exit. This means that very poorly run government-owned companies can remain in business for an almost indefinite period of time without earning sufficient revenues to cover their costs. This can induce severe distortions in the allocation of production across firms in the industry. A very inefficient government-owned firm can continue to serve a large fraction of demand, much to the detriment of more efficient suppliers. This logic provides another rationale for ending government participation in these sectors of the industry as soon as it is feasible. It also underscores the necessity of shoring up a country's bankruptcy laws before going forward with the re-structuring process.

5. Lessons for Market Design from Specific LACs

This section discusses the market design challenges specific to each of the five LACs that I visited. Some of these challenges were discussed in general terms in the previous section, whereas others are unique to the country under consideration. The primary goal of this section is to point out how the market designer must adapt the ESI restructuring process to the initial conditions in the industry each country. The quantity and mix of generation capacity, the level and annual pattern of demand, the legal framework governing the industry and geographic distribution of demand must all be accounted for in the market design process.

5.1. Brazil

Brazil is the prototypical example of the gambling with the weather problem discussed in the Section 3. Hydroelectric facilities produce the vast majority of electricity. The cost-of-deficit parameter used in the cost-based stochastic dynamic programming dispatch model is extremely low relative to international standards, and the recommendations given in Section 3 apply directly to Brazil. The cost-of-deficit parameter should be increased to levels that reflect the social cost of deficits and retail rates should be adjusted on a seasonal basis to reflect current conditions in the wholesale market. Retail prices that reflect current water availability in the retail price will ensure that periods of shortage do not occur.

Another significant source of inefficiencies in the current Brazilian market design is the Energy Reallocation Mechanism (MRE). This mechanism pays hydroelectric suppliers according to the share of assured energy certificates (CEAs) they own rather than according to their actual hourly output. CEAs are awarded to each hydroelectric facility in Brazil according to a somewhat arbitrary administrative process. Each hour the total quantity of hydroelectric energy produced in Brazil is measured, and this production is allocated to each hydroelectric supplier according to the share of the total system-wide CEAs allocated to this generation unit, regardless of how much energy the unit supplies during the hour. In contrast, fossil fuel units are paid the hourly spot price based on their actual production.

The MRE distorts plant investment decisions across fuel types and locations in the transmission network. A new hydroelectric entrant earns revenues based on the number of CEAs it obtains from the administrative process for constructing the unit, not based on the actual hourly output of the unit. Consequently, one can easily imagine a circumstance where the hydroelectric unit that provides the greatest benefits to system reliability is not built because the administrative process for construction awards a larger quantity of CEAs to a unit at a different location in the transmission network. The MRE also induces a bias against the construction of fossil fuel facilities, because they must factor in the risk of not being dispatched into their revenue risk calculation, whereas hydroelectric facilities are guaranteed their share of the system-wide output of hydroelectric facilities whether or not they produce any electricity during the hour.

The MRE hinders the development of liquid forward market for electricity because it significantly reduces the incentives for hydroelectric suppliers to be active participants in this market, as they already have a relatively certain revenue steam for their facilities. Because the MRE pays hydroelectric generators whether their unit operates or not, this significantly dulls the incentives for owners to maintain their units in top working order. A major benefit of the spot market default payment mechanism in all other electricity markets is that it pays suppliers the highest prices during the hours in which their units are most needed to maintain system reliability. Consequently, unit owners have a strong incentive to maintain their units in top working order in case an hour or series of hours arises when the units will receive a particularly high price for supplying electricity.

The MRE should be phased out as soon as possible. The longer it is allowed to remain in force, the greater will be the amount of new capacity built at the wrong location in the transmission network and using the wrong input fuel. The most straightforward way to accomplish this is by a multi-year scheme that reduces the amount of MRE-based payments each year. For example, during the first year, 90 percent of the revenues could come from the MRE and the remaining 10 percent from spot market production and sales. Over time the MRE share would fall and the spot market portion rise, until at the end of a pre-specified time period the MRE share would equal zero. Similar mechanisms are used to phase out vesting contracts associated with generating facilities sold by the government at the start of the restructuring process.

Brazil has also had substantial difficulties establishing a credible regulatory mechanism. The rationing period in 2001 severely undermined the public's and government's confidence in the regulatory body, ANEEL. This is particularly unfortunate for Brazil, where the major load centers are located far from the new sources of hydroelectric power. Expansion of the transmission network is therefore essential to meeting Brazil's demand growth with new hydroelectric capacity. A strong regulator is needed to oversee this transmission expansion process. Also, and as quickly as possible, Brazil needs to develop the necessary regulatory expertise and credibility to determine the magnitude and location of transmission upgrades. It is very unlikely that the necessary new transmission capacity will be built without the promise of a regulated rate of return on the investment, so the regulator must also devise a mechanism for sharing the burden of paying for these transmission investments among market participants; managing these processes would strain the regulatory process even in most developed countries. Therefore, Brazil appears to be an ideal candidate for the independent international committee of experts described in the previous section. Such a committee could provide much need expertise and credibility to the regulatory process.

Finally, Brazil is currently considering a single-buyer model for wholesale electricity procurement. For the reasons listed in the previous section, this is a step backwards to the former state-owned, vertically integrated monopoly regime rather than a step forward to the wholesale competition that is likely to benefit Brazilian consumers.

5.2. Chile

Chile has the most mature electricity market in Latin America. It has been in operation for almost twenty years, and there are four separate networks in Chile. The Central Interconnected System (SIC) is by far the largest. It serves 90 percent of the population and more than 40 percent of the land area. The majority of the generation capacity in the SIC is hydroelectric. The Great North Interconnected System (SING) serves primarily mining consumers and is almost entirely thermal. The other two grids represent only a small portion of the installed capacity in Chile.

The designers of the Chilean market recognized the importance of developing an active forward market for energy. Since the initial ESI reforms, the SIC has operated a cost-of-service spot market using a series of stochastic dynamic programming models to determine the opportunity cost of water from the variable cost of fossil fuel units and a cost-of-deficit parameter. The model has led to a substantial amount of new generation capacity constructed in both the SIC and SING. Most of this capacity uses natural gas and coal. The availability of natural gas from Argentina has allowed the construction of a number of combined-cycle gas turbine facilities. Viewed from the perspective of ensuring capacity adequacy, the Chilean market has been successful. However, for many of the same reasons as Brazil, it has had problems ensuring energy adequacy in the short run in response to low hydro conditions.

During the period 1998-1999, Chile had an energy shortage because of a sustained period of low water availability coupled with a retail price system that does not reflect current conditions in the wholesale market in retail electricity rates. Fischer and Galetovic (2001) also attribute these electricity shortages to a regulatory mechanism that failed to act quickly and aggressively enough to prevent the shortage period from occurring. While the severity and duration of these shortages may have been enhanced by the shortcomings of the regulatory governance mechanism, the cause of the shortage appears to be the gambling with the weather problem described in Section 4.1. The cost-of-deficit parameter in the stochastic dynamic programming models used to dispatch hydroelectric capacity is far below the highest willingness to pay for electricity of all consumers in Chile. In fact, it is of the same \$/MWh order of magnitude as value that exists in Brazil at the present time.

The water level in the Laja reservoir, the largest reservoir in Chile, determines the opportunity cost of water. When full, this reservoir holds enough water to generate about one-quarter of Chile's annual consumption of electricity. The stochastic dynamic program used to set the opportunity cost of water trades off the benefit of using water today against the cost of using water in the future and therefore having to operate thermal capacity or ration energy. As discussed in Section 4.1, a low cost-of-deficit parameter tells the stochastic dynamic program that water shortages are not too costly, and as consequence, water will be used more intensively than it should to safeguard against shortage periods. Spot wholesale electricity prices are also lower than they should be, which encourages over-consumption of electricity and increases the likelihood of shortage periods.

As discussed in Section 4.1, during most years the gambling with the weather strategy will be successful because sufficient water is available to meet the system's energy needs. However, during low water conditions, the hydro system can reach a point-of-no-return from a shortage period. By this I mean that at the water levels in the reservoirs at that time, even if the thermal plants are operated as much as they are physically able, there will be an annual energy shortfall. As discussed in Section 4, during these circumstances, retail prices must be raised to cause demand to ration the available supply. Although this eventuality cannot be avoided with certainty by raising the cost-of-deficit parameter, it can be reduced to an extremely small probability by a sufficiently high cost-of-deficit parameter. This higher cost-of-deficit parameter will cause fossil fuel plants to be operated more frequently in order to use the available water more conservatively. The higher this cost-of-deficit parameter is set the lower will be the probability of a period when retail prices must be increased to reduce demand to the level of the available supply.

The geography of Chile increases the share transmission network costs in the delivered price of electricity relative other countries in Latin America and around the world. The country is extremely long and thin, so that transmission losses can be very large for demand located far from the generation centers. One component of the revenue stream to the transmission owner is associated with the fact that the marginal cost of withdrawals in a generation-rich area are lower than the marginal cost of withdrawals in a generation-deficient area. However, these payments are not in general sufficient to compensate the transmission owner for all of its costs. This has led the regulator to implement a toll component that every generation unit owner must pay to the transmission owner. This is a fixed charge based on a number of administrative procedures. Specifically, the regulator determines the "Area of Influence" of a generator, which is the minimum combination of lines, substations and other transmission network installations that allow each generator to connect to the reference node of the electricity network. Each generator is then assessed a transmission charge in proportion to their "Area of Influence." There is an additional toll component to account for energy produced by the generator that is carried further than the generator's area of influence.

These two procedures for assessing transmission tolls are viewed by many generation unit owners as arbitrary. These concerns seem valid because it is difficult to see how these cost-allocation procedures can be rationalized based on economic cost causation principles. Because all generation units are connected to the same transmission network, and because the laws of physics rather than economic contract paths determine energy flows in the network, it is extremely difficult to allocate all transmission network costs on the basis of cost-causation principles. In particular, all electricity produced by all units in the SIC is not delivered to the reference node for the electricity system.

Any mechanism to pay for transmission network should allocate all costs that are caused by the actions taken by specific market participants to the market participant that cause them. However, the remaining costs that cannot be causally allocated should be clearly identified and those should be recovered from market participants in a manner that induces the least costly distortions in generation unit operating decisions and new plant location decisions. Cost-allocation schemes based on observable characteristics on market participants, such as where they are located in the transmission network, can cause market participants to take actions to reduce the costs that are allocated to them. These actions may also degrade system's reliability or increase wholesale energy prices. To avoid these problems, all of the ISOs in the United States recover the non-causal portion of costs of the transmission network from a per MWh charge on all withdrawals of energy from the transmission network.

Currently the SIC and SING are not interconnected, in spite of the fact that prices in SING are extremely low because of a large amount of generation capacity in the northern part of Chile. A transmission cost allocation scheme, which makes a clearer distinction between causal costs that can be attributed to specific market participants and common costs that must be recovered from all market participants, may make it easier to construct an interconnection between these two systems. This would provide substantial reliability benefits for the hydro-dominated SIC and access for the SING to the SIC's substantial hydroelectric resources.

Chile also provides a concrete example of a major conclusion from Section 2 concerning the importance of accounting for the individual rationality constraint in the market design process. Chile has implemented what in theory appears to be an extremely high-powered incentive regulation scheme for its distribution utilities. The electricity law in Chile requires the regulator to construct an "efficient distribution company" from which to derive the rates that will apply to all companies of that type, where the type of company is measured by the density of customers served in the service territory. In theory, this scheme should provide very strong incentives for companies to reduce their production costs, because their regulated price does not depend on their own actions. However, as Alejandro Jadresic (the Minister of Energy from 1994-1998) notes, the threat of judicial review forces the regulator to look at actual company performance. Dyck and DiTella (2002) analyze the performance of this regulatory mechanism and reproduce the following quote from Jadresic, “When building the model, you end up always looking at what happens in actual companies. The cost studies are subject to the challenge of verifiability in court; the model company could be regarded as a pure ungrounded imaginary construction.”

Dyck and Di Tella (2002) provide convincing empirical evidence that companies also recognize this fact and appear to alter their balance sheet costs in preparation for a coming regulatory review. Although Dyck and Di Tella find significant cost reductions among distribution companies over the course of each price cap period, they find a U-shaped pattern to these cost reductions. Trends in cost reductions are reversed every four years, apparently in response to the regulatory reviews that set a new price cap, which also occur every four years. Consistent with the predictions implied by the individually rationality constraint on distribution company behavior and the fact that distribution companies recognize the regulator must account for their current financial position or risk judicial review, the cost reductions achieved during the initial years of the price cap regime slow down and reverse as the time of a new review approaches.

Dyck and DiTella’s results should not be taken as an argument for the inferiority of price cap regulation versus other forms of regulation. Instead, it should be seen only as an argument in favor of the position that designing a regulation process that achieves productive efficiency is impossible, and that individual rationality implies that firms will maximize their objective function subject to the constraints on their behavior implied by the regulatory process.

5.3. Colombia

Columbia is one of the few LACs that uses a bid-based spot market. Surprisingly this did not prevent shortage periods due to lack of water availability from occurring in Colombia. This provides empirical evidence that a bid-based market will not prevent apparent shortage periods from occurring. Colombia faces the same problem other hydro-based system in Latin America face, the inability to reflect current conditions in the wholesale market in the retail rates that consumers pay. This means that apparent shortage periods will occur when the annual demand at the prevailing retail price exceeds the amount of available water assuming that the available fossil fuel units are operated as intensively as possible.

There is an additional reason for Colombia to adopt a cost-based dispatch market rather than a bid-based one. The transmission network is subject to frequent de-ratings, and many of the hydroelectric facilities are located far from the load centers. This means that opportunities for suppliers to exercise local market power because of transmission line outages can be significant. For this reason, a cost-based dispatch, in spite of the problems mentioned in Section 4.4, seems to make better sense for Colombia.

Colombia has also had difficulty establishing a credible regulatory mechanism. This has been exacerbated by the existence of a bid-based spot market that requires the regulator to determine when suppliers' exercise of unilateral market power harms system reliability and market efficiency enough to justify regulatory intervention. Formulating a regulatory mechanism to deal with this problem has been extremely challenging for developed countries such as the United States and the United Kingdom, so it not surprising that the Colombian regulator has found this difficult. This fact is another argument against a bid-based spot market in Colombia. If a cost-based spot market is adopted, I expect the regulatory challenges described in Section 4.4 to arise. For these reasons, Colombia appears to be another ideal candidate for the independent advisory committee of experts to assist the regulator establishing the necessary expertise and credibility.

5.3. Honduras

Honduras is representative of the challenges faced by a number of small LACs. It has a peak demand of approximately 800 MW and an installed capacity of slightly over 900 MW. Slightly less than half of the installed capacity is hydroelectric. The remaining capacity uses fossil fuels, although Honduras has no significant domestic fossil fuel energy sources. Electricity demand is expected to grow rapidly for the near future, primarily because a significant fraction of the population does not currently have access to electricity.

Restructuring in a country as small as Honduras is complicated by the fact that the minimum efficient scale of a new combined cycle natural gas facility is a substantial fraction of the peak demand in the country. For this reason, anything but a cost-based spot market is out of question for Honduras. It should concentrate on developing an active forward market where privately-owned companies can sell the necessary forward energy commitments necessary to build the new capacity needed to meet the country's growing demand. Because Honduras has no significant domestic fossil fuel supplies, there are clear benefits to encouraging diversity in the input fossil fuels that its plants can burn.

The potential benefits from expanded transmission interconnections with other Central American countries are substantial. This should be a top priority for the restructuring process in Central America. With expanded interconnections, Central American countries can increase the capacity utilization rates on base load fossil-fuel facilities and share plentiful water conditions or more efficiently manage low water conditions throughout the region. Central American countries could also maintain the same level of domestic grid reliability with less capacity in the control area, because capacity outside the region could be relied upon to provide operating reserves.

Increased regional coordination among the country-level regulators in Central American countries would increase the likelihood that these interconnections would be built. The independent committee of experts approach once again appears to be a viable solution to this problem. Establishing such a committee for the entire Central American region would spread the cost over a number of countries and foster information-sharing across the various countries. One role of the committee would be to facilitate the coordination of country-specific regulatory processes.

5.4. Mexico

Mexico has the potential to realize substantial benefits from ESI restructuring. It has abundant fossil fuel energy sources, particularly oil and natural gas, and average retail electricity prices seem high relative to international standards given this relatively abundant natural resource base. Mexico continues to experience rapid demand growth fueled by a robust national economy and expanding industrial sector. A major concern of the Mexican government is the ability of the state-owned monopoly to construct the necessary generation capacity to meet this demand growth. Unfortunately, in spite of the clear need for ESI restructuring and the large potential benefits to the citizens of Mexico, the necessary capacity expansion is unlikely to occur in the near future.

In this regard, Mexico is a prime example of the necessity of accounting for initial conditions in the country and industry in formulating a restructuring strategy. The Mexican constitution explicitly prohibits any entity but the government from providing electricity for public service. Unless this provision is amended the options available to Mexico for restructuring its industry are extremely limited.

However, there has been one benefit to this legal barrier to industry restructuring. The energy regulatory commission in Mexico (CRE) has been studying this issue for the past two years. In the process, it has interacted extensively with international experts and gained considerable expertise in electricity market design, which should enhance its credibility in the event that ESI restructuring eventually takes place.

6. Suggested Standard Market Design For LACs

This section builds on the analysis of the previous sections to outline the general features of a standard market design for LACs. I also note how this standard design can be modified as more of the necessary conditions for a competitive wholesale market are realized in each country.

The first step in any market design process is to establish an independent regulator or regulatory body. The US model of a regulatory body appointed by the executive branch of the government subject to the approval of the legislative branch is preferable. Decisions should be made based on a legal record prepared by the parties involved, managed by an administrative law judge and supported by a regulatory staff of economic, legal, and power systems engineering experts. The staff of the regulatory body should be hired in anticipation of the duties they will perform, rather than in response to the demand for additional regulatory oversight.

As discussed is previous sections, the regulatory process must be forward looking, rather than backward looking. Market design flaws and other regulatory problems must be identified and addressed as quickly as possible. There should be a substantial training component at the start of the market design process. For example, visits by the regulatory commission staff to other functioning wholesale markets around the world would provide useful background. Failure to learn from international experience can be extremely costly. Many of the early market failures in the US could have been avoided had FERC taken more seriously the lessons from other markets around the world. Specifically, these markets had substantial experience dealing with system-wide and local market power issues that FERC, for the most part, ignored until the aftermath of the California crisis.

Once the regulatory process is in place, the first task for this entity is to develop a forward market for electricity where private investors sell obligations to supply electricity that can be used to finance new generation capacity. At least for the foreseeable future, establishing a formal bid-based spot market seems too costly in terms of potential consumer harm to justify. Problems with unilateral and local market power have proven extremely difficult for developed countries to solve, and many of them have a long history with regulation and competition policy that most LACs do not share

Because all suppliers need to buy and sell deviations from their final day-ahead schedules or longer-term energy schedules, a real-time market is necessary. This can be accomplished by the former vertically integrated monopolist operating a real-time imbalance market using cost-based bids. All suppliers must file their costs with the ISO and regulator, and after these costs are validated by the regulator they are made publicly available to all market participants. The ISO then dispatches all units based on these costs, which also produces locational marginal prices (LMPs) at all nodes in the network. It is not essential that suppliers be paid or pay their LMP for deviations from their final energy schedules. Retailers and large consumers could also be charged prices aggregated over larger geographic areas. However, LMPs are the best signals of the cost of withdrawing power at each node in the network, so the regulator should have a very good reason for not paying generation units according to these prices and charging load-serving entities for their wholesale purchases at these prices.

Initially there is little need to divest capacity from the incumbent monopolist. It is more important for the regulator to focus on obtaining credible start-up, no-load and variable cost figures for all units in the control area. The dispatch process should also incorporate a more realistic cost of shortage parameter. The easiest way to set the cost-of-deficit parameter is to simulate hydro conditions and fossil-fuel unit variable costs for a number of values of this parameter to determine the relationship between the probability of shortages and the value of the cost-of-deficit parameter. The regulator should then set the initial value of this parameter to achieve the level and probability of shortage less than a value jointly agreed upon by all market participants.

The goal of cost-based dispatch for imbalances in real-time is to establish a transparent mechanism that all market participants can use to assess the costs and benefits of using this imbalance mechanism. New entrants can factor these costs into their willingness to supply energy on long-term contracts at specific locations in the transmission network. Cost-based dispatch also avoids most of the problems associated with a transmission network that cannot support a competitive wholesale spot market, an initial condition that exists in almost all LACs. Setting LMPs using cost-based bids will provide useful information to the ISO about the benefits of transmission upgrades in the network and important input into the long-term process of constructing an economically reliable transmission network.

A cost-based dispatch mechanism also allows the ISO to be a market participant. However, because the ISO must run a cost-based dispatch based on publicly available cost data, there is less concern with permitting the ISO to take a position in the real-time dispatch.

Once this dispatch process has been established, the process of opening the wholesale market to consumers can begin. This should be demand driven. By this I mean that to the extent that large consumers are willing to subject themselves to the hourly spot price as their default price, the wholesale market should grow.

This market structure implies two types of consumers. The first are negawatt suppliers, who are the demand-side equivalent of privately owned generation owners. They must purchase all of their demand at either the hourly spot price or at a forward contract price they have managed to negotiate with some electricity supplier. The second are those who wish to remain with their monopoly retailer. The monopoly retailer for their geographic area must manage the spot price risk associated with serving these captive customers. The regulator should encourage the default suppliers to set retail prices that vary with system conditions for captive consumers.

The difference between the negawatt suppliers and captive customers is that the former group can shop around with any supplier for a better forward contract price for their electricity needs but can never return to being a captive consumer. Because the negawatt suppliers cannot return to their default provider, in exchange for the opportunity to pay a lower price, they now face the risk that there will not be enough new capacity to meet their demand. This will give them incentives to enter into forward contracts that can be used to finance new investments.

In order to set the retail price the monopoly retailers must pay for wholesale electricity, the regulator will run periodic auctions for standardized contracts for electricity supply and retailers will be required to buy a pre-specified fraction of their load obligations each year in the future from these markets.

A final issue with this market design is how to handle the issue of system-wide shortages of water. The only way to ultimately solve this problem is to raise the cost-of-deficit parameter to a high enough level to reflect the full economic and political costs of shortages. Because of the potential for a substantial retail price increase if the cost-of-deficit parameter is increased, this parameter should be slowly increased over time. At the same time, the transition should begin to expose all the regulated consumers to retail prices that vary with actual system conditions. In this way, the theoretical ideal cost-of-deficit parameter can be set, and can consumers decide the retail price at which they are willing to curtail their demand. This cost-of-deficit parameter, combined with pricing to final consumers that reflect current conditions in the wholesale market, will completely eliminate the possibility of system-wide shortages.

7. Concluding Comments on Designing Competitive Wholesale Markets in LACs

Wholesale markets in LACs should be designed for the initial conditions that exist in the ESI. The successful market designs in developed countries and LACs emphasize this very important lesson. The California electricity crisis provides a graphic example of the economic harm that is possible when a restructuring process ignores the initial conditions in the market. In California's case, coupling a wholesale market where electricity generation unit owners and energy traders had enormous flexibility to take action to maximize their profits with a retail market that put substantial restrictions on the actions of the large retailers created a recipe for disaster. Combining this with a regulatory process poorly prepared to deal with the regulatory problems that can arise in a wholesale market environment, allowed a flawed market design to turn into an economic disaster from which California is still attempting to recover.

Consequently, the goal for ESI reform in LACs is gradual progress to the ultimate goal of a competitive market that benefits consumers with the appropriate safeguards in place at the start of the market to protect them from significant financial harm. Another important reason for this continuous improvement strategy is that it allows a nascent regulatory process to gain expertise and credibility. With substantial safeguards in place at the start of market, the regulator's job is much more straightforward. However, as these safeguards are released, her job becomes far more challenging.

Finally, even though a slow but steady strategy for ESI reform seems optimal for LACs, it is still important to make this process consistent with achieving the market designer's long-term goals of a competitive wholesale market that yields the maximum benefits possible to consumers. Therefore, the market designer should avoid locking-in beneficial safeguards that are extremely difficult to remove and are also very costly to long-term market efficiency.

References

Averch, H., and L. Johnson. 1962. "Behavior of the Firm Under Regulatory Constraint." *American Economic Review* 52: 1053-1069.

Araújo, J.L. 2002. "Investment in the Brazilian ESI - what went wrong? What should be done?, Rio de Janeiro, Brazil: Instituto de Economia da Universidade Federal do Rio de Janeiro.

Ayala, U., and J. Millan. 2002. "The Sustainability of Power Sector Reform in Latin America: The Reform in Colombia." Paper prepared for the Inter-American Development Bank. http://216.239.57.100/search?q=cache:kNCah5qfJ3AJ:www.iadb.org/sds/doc/Workshop_material-2-E.pdf+ayala+millan+colombian+power+sector&hl=en&ie=UTF-8

Borenstein, S., J. Bushnell F.A. Wolak. 2002. "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market." *American Economic Review* 92(5): 1376-1405.

Borenstein, S. 2003. "Time-Varying Retail Electricity Prices: Theory and Practice." Paper presented at 8th Annual Research Conference on Electricity Industry Restructuring, March 14, 2003. Berkeley, United States: University of California, Department of Economics.

Brown, A.C. 2002. "The Privatization of Brazil's Electricity Industry: Sector Reform or Restatement of the Government's Balance Sheet?" Washington, DC, United States: Inter-American Development Bank. http://www.iadb.org/res/index.cfm?fuseaction=Publications.View&pub_id=S-141

Bushnell, J., and F.A. Wolak. 1999. "Regulation and the Leverage of Local Market Power in the California Electricity Market." http://www.stanford.edu/~wolak.

Bushnell, J., E. Mansur and C. Saravia. 2002. "An Empirical Assessment of the Competitiveness of the New England Electricity Market." Center for the Study of Energy Markets (CSEM) Working Paper 101. http://www.ucei.berkeley.edu/PDF/csemwp101.pdf

Dyck, I.J. Alexander, and Di Tella, Rafael. 2002. "Cost Reductions, Cost Padding and Stock Market Prices: The Chilean Experience with Price Cap Regulation." Harvard Business School NOM Research Paper No. 03-22. Cambridge, United States: Harvard University.

Fischer, R., and P. Serra. 2000. "Regulating the Electricity Sector in Latin America." Serie

Economía No. 86. Santiago, Chile: Universidad de Chile.

Fischer, R., and A. Galetovic. 2001. "Regulatory Governance and Chile's 1998-1999 Electricity Shortage." World Bank Policy Research Working Paper 2704. Washington, DC, United States: World Bank.

Galal, A., L. Jones, P. Tandon and I. Vogelsang. 1994. *Welfare Consequences of Selling Public Enterprises: An Empirical Analysis*. Oxford, United Kingdom: University Press.

Green, R., and T. McDaniel. 1998. "Competition in Electricity Supply: Will '1998' Be worth It?" POWER working paper. PWP-057. http://www.ucei.berkeley.edu/PDF/pwp057.pdf

Green, R.J. 1999. "The Electricity Contract Market in England and Wales." *Journal of Industrial Economics* 47(1): 107-124

Honduras-Presidential Commission for State Modernization. 2002. "Design of Power Sector Reform, October.

Joskow, P., and E. Kahn. 2002. "A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000: The Final Word." *Energy Journal* 23(4): 1-36.

Kolstad, J.T., and F.A. Wolak. 2003. "Using Environmental Emissions Permit Prices to Raise Electricity Prices: Evidence from the California Electricity Market." http://www.stanford.edu/~wolak.

Mexico, Ministry of Energy. 2002. "Proposal for the Electrical Sector Modernization." Mexico City, Mexico: Ministry of Energy.

Millan, J., E. Lora and A. Micco, Alejandro. 2001. "Sustainability of Electricity Sector Reforms in Latin America." Prepared for Seminar, Towards Competitiveness: The Institutional Path, March 16, 2001, Santiago, Chile, Inter-American Development Bank. http://www.iadb.org/res/index.cfm?fuseaction=Publications.View&pub_id=S-141

Newbery, D. 2002. "Issues and Options for Restructuring Electricity Supply Industries." Department of Applied Economics Working Paper WP 0210. Cambridge, United Kingdom: Cambridge University, Department of Economics.

Parente, P. 2002. "Looking beyond the Energy Crisis: Scenarios for a Renewed Brazilian Electric Power Sector." Address at the Conference on Energy in South America, Center for Latin-American Issues, George Washington University, Washington, DC, United

States. http://www.energiabrasil.gov.br/pronunciamentos/pronunc_parente_3.asp

Patrick, R.H., and F.A. Wolak. 1997. "The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market." http://www.stanford.edu/~wolak.

Power Systems Research, Inc. (PSRI). 2002. "Brazilian Power Sector Reform." Rio de Janeiro, Brazil: PRSI.

Rufin, C. 2002. "Political Economy of the Brazilian Electricity Sector." Report presented to the Inter-American Development Bank. Washington, DC, United States, Inter-American Development Bank.

Spiller, P. T., and L.V. Martorel. 1996. "How Should It Be Done? Electricity Regulation in Argentina, Brazil, Uruguay, and Chile." In: R.J. Gilbert and E. Kahn, editors. *International Comparisons of Electricity Regulation*. Cambridge, United Kingdom: Cambridge University Press.

Walker, I., and J. Benavides. 2002. "Sustainability of Power Sector Reform in Latin America: The Reform in Honduras." Washington, DC, United States: Inter-American Development Bank, Sustainable Development Department. http://www.iadb.org/sds/doc/IFM-Case-Homduras-E.pdf

von der Fehr, N-H. and F.A. Wolak. 2003. "Report on Brazilian Power Sector Reform." http://www.stanford.edu/~wolak.

Wolak, F.A. 1994. "An Econometric Analysis of the Asymmetric Information Regulator-Utility Interaction." *Annales d'Economie et de Statistique* 34: 13-69.

----. 1999. "Market Design and Price Behavior in Restructured Electricity Markets: An International Comparison." In: T. Ito and A. Krueger, editors. *Competition Policy in the Asia Pacific Region*. NBER East Asian Seminars on Economics Volume 8. Chicago, United States: University of Chicago Press.

----. 2000. "An Empirical Analysis of the Impact of Hedge Contracts on Bidding Behavior in a Competitive Electricity Market." *International Economic Journal* 14(2): 1-40.

----. 2001. "Designing a Competitive Electricity Market that Benefits Consumers." http://www.stanford.edu/~wolak

----. 2002. "Competition-Enhancing Local Market Power Mitigation in Wholesale Electricity

Markets." http://www.stanford.edu/~wolak.

----. 2003a. "Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Electricity." In: M. Dewatripont, L.P. Hansen and Stephen J. Turnovsky (editors). *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*. Volume II. New York, United States: Cambridge University Press.

----. 2003b. "Measuring Unilateral Market Power in Wholesale Electricity Markets: The California Market 1998 to 2000." Forthcoming in *American Economic Review*. http://www.stanford.edu/~wolak.

----. 2003c. "Diagnosing the California Electricity Crisis." http://www.stanford.edu/~wolak.

Wolak, F.A., R. Nordhaus and C. Shapiro, Carl. 2000. "Analysis of 'Order Proposing Remedies for California Wholesale Electric Markets (Issued November 1, 2000).'" Market Surveillance Committee of the California Independent System Operator. http://www.caiso.com/docs/2000/12/01/2000120116120227219.pdf

Figure 1. Residual Demand Elasticity and Profit-Maximizing Behavior

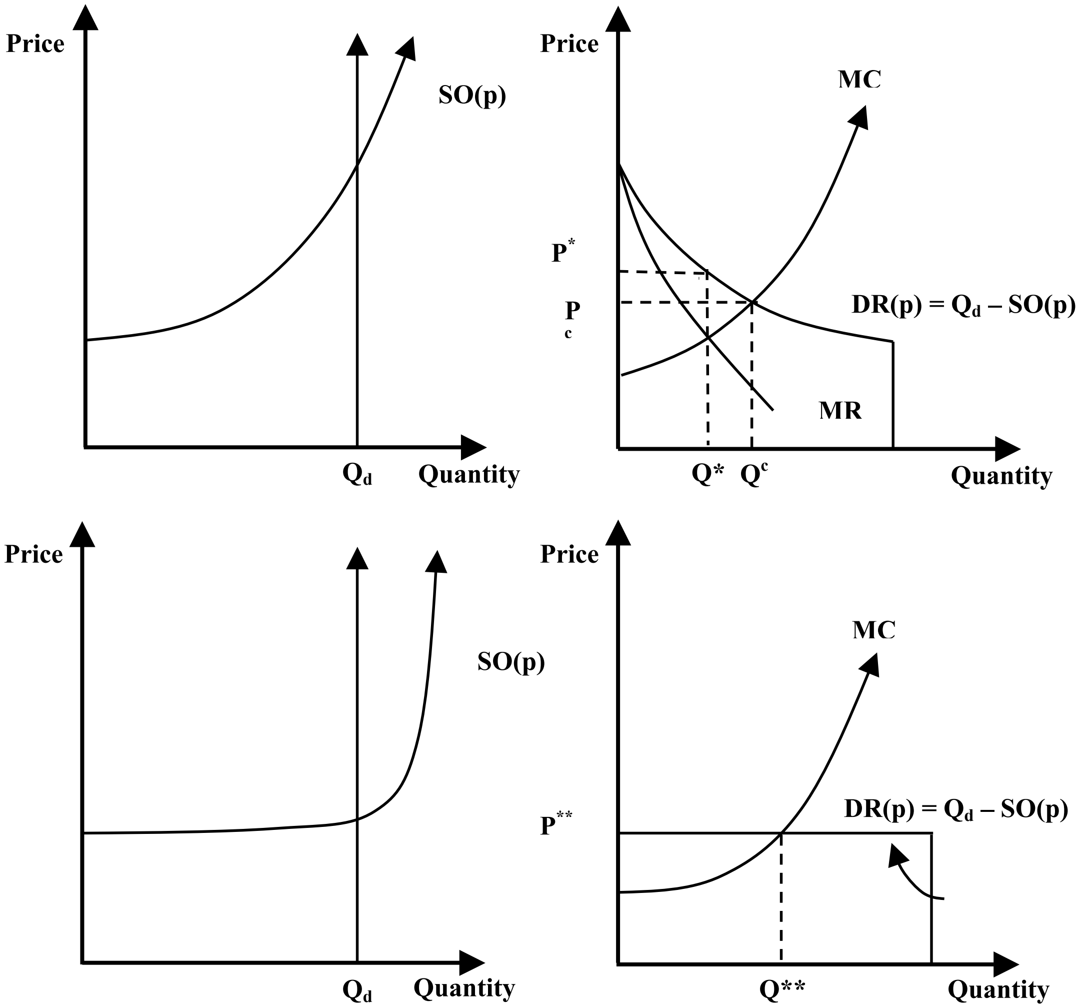

Figure 2. Welfare Loss from Inefficient Production

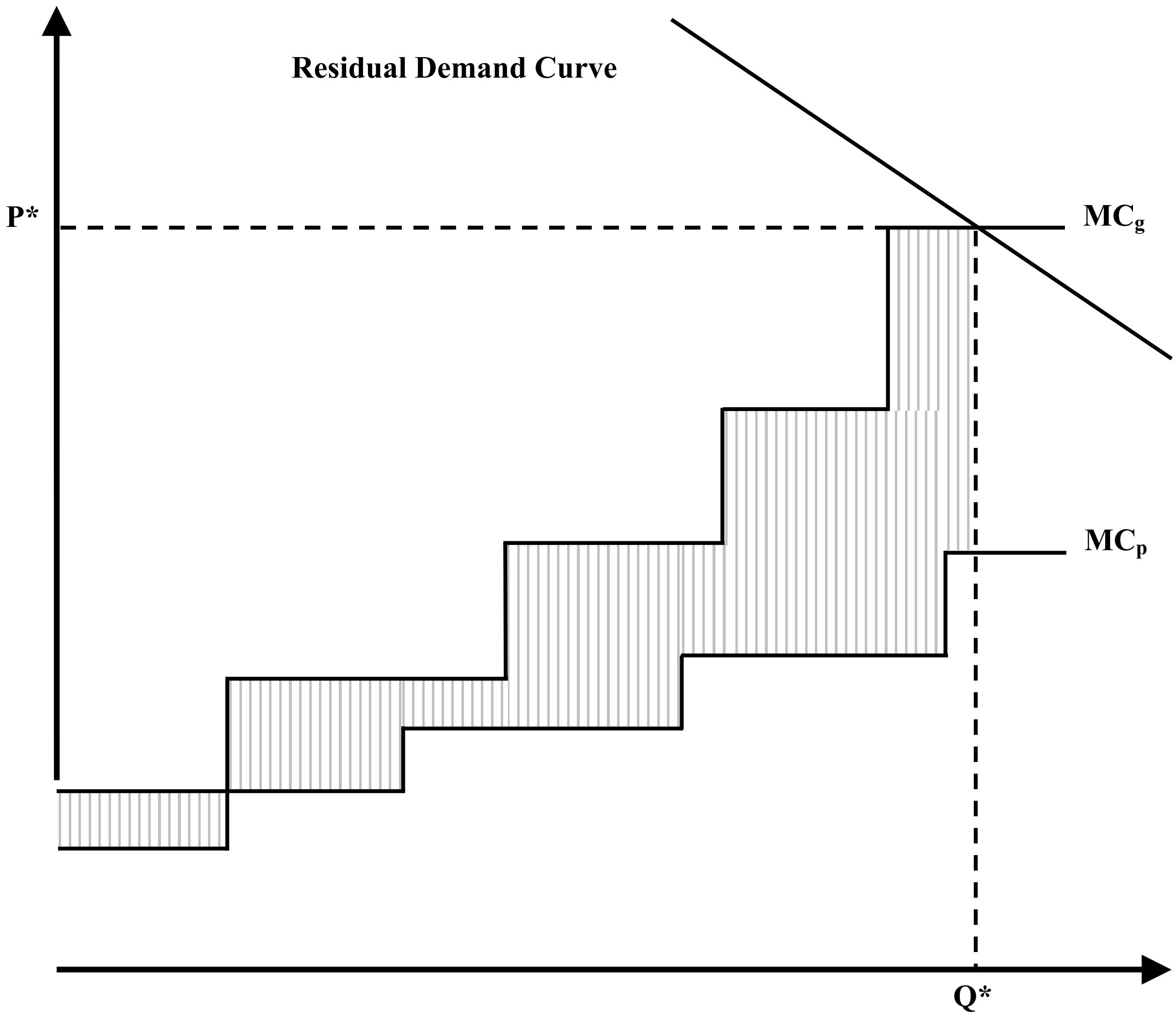

Figure 3. The Impact of Capacity Divestiture on a Pivotal Supplier

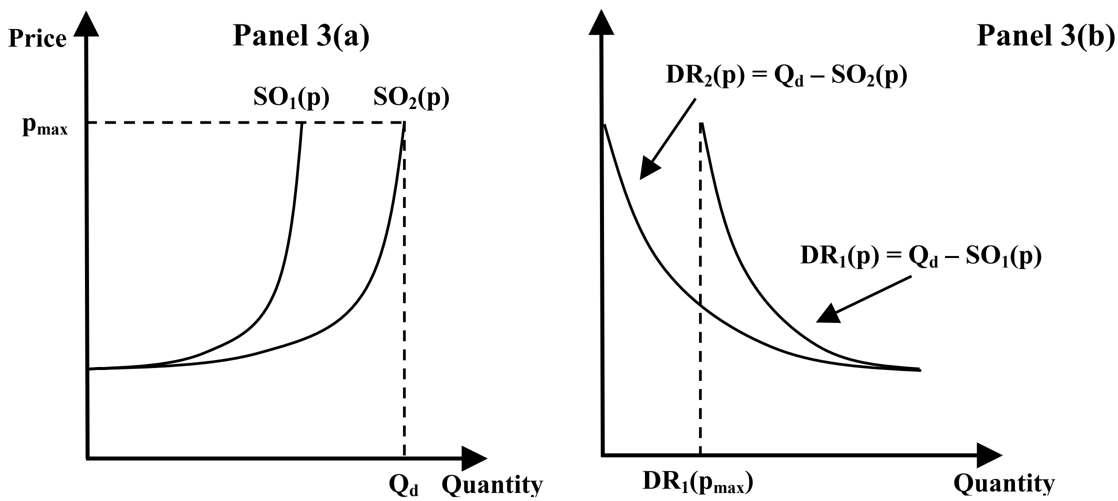

Figure 4. Residual Demand Elasticity and Price-Responsive Demand

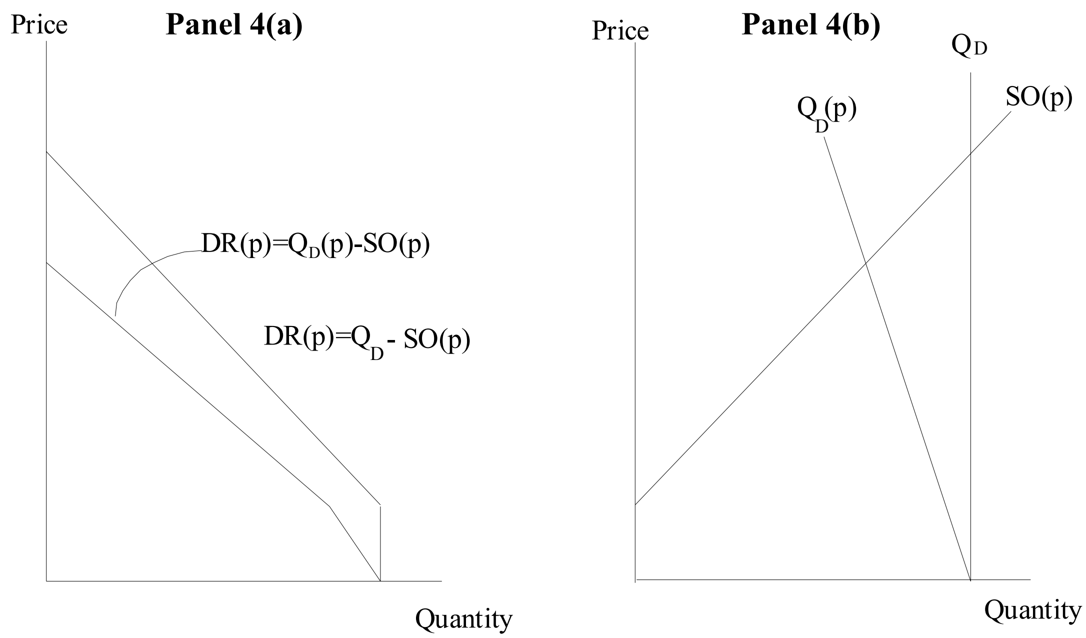

Figure 5. Retail Pricing, Hydro Availability and Supply Shortfalls